GenCore version 5.1.4_p5_4578

OM protein - protein search, using sw model

Run on: April 22, 2003, 15:28:07 ; Search time 28 Seconds
(without alignments)
1482.779 Million cell updates/sec

Title: US-10-046-433-40
Perfect score: 1001
Sequence: 1 MAEPGHSHHLSARVRGRTER.....LGRSNHLPPRGLLMDLTQCR 1001

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8	0.8	284	1	KIME_ARCFU	O27995 archaeoglob
2	8	0.8	353	1	PERA_ARMRU	P00433 armoracia r
3	8	0.8	354	1	PERC_ARATH	P24101 arabidopsis
4	8	0.8	443	1	TOLB_RICPR	Q9zdm5 rickettsia
5	8	0.8	4644	1	DYHC_MOUSE	Q9jhu4 mus musculu
6	7	0.7	73	1	NER_BPD10	P06903 bacteriopha
7	7	0.7	75	1	LSM6_SCHPO	Q9uui1 schizosacch
8	7	0.7	80	1	LSM6_HUMAN	Q9y4y8 homo sapien
9	7	0.7	100	1	URE3_ECOLI	Q03282 escherichia
10	7	0.7	100	1	URE3_PROMI	P17088 proteus mir
11	7	0.7	100	1	URE3_PROVU	P16124 proteus vul
12	7	0.7	175	1	DEF_RICCN	Q92iz1 rickettsia
13	7	0.7	175	1	Y773_METJA	Q58183 methanococc
14	7	0.7	181	1	YCF4_ODOSI	P49526 odontella s
15	7	0.7	215	1	PMG2_ECOLI	P36942 escherichia
16	7	0.7	239	1	DHSB_RECAM	P80480 reclinomona
17	7	0.7	244	1	IF6_SCHPO	O94476 schizosacch
18	7	0.7	267	1	TRY7_ANOGA	P35041 anopheles g
19	7	0.7	292	1	YC15_PYRHO	O58968 pyrococcus
20	7	0.7	293	1	RRP4_HUMAN	Q13868 homo sapien
21	7	0.7	300	1	DPPC_ECOLI	P37315 escherichia
22	7	0.7	320	1	STBA_ECOLI	P11904 escherichia
23	7	0.7	346	1	YA0D_SCHPO	Q09681 schizosacch
24	7	0.7	350	1	AROG_ECOLI	P00886 escherichia
25	7	0.7	362	1	AROG_HAEIN	P44303 haemophilus
26	7	0.7	439	1	AKR_ARATH	Q05753 arabidopsis
27	7	0.7	450	1	CBPA_SYNP7	P39660 synechococc
28	7	0.7	451	1	TBA2_NEUCR	P38669 neurospora
29	7	0.7	457	1	RADA_LISMO	Q48761 listeria mo
30	7	0.7	482	1	CAB1_METJA	Q58773 methanococc
31	7	0.7	515	1	PEN3_ADEG1	Q64755 avian adeno
32	7	0.7	538	1	RO60_HUMAN	P10155 homo sapien
33	7	0.7	538	1	RO60_MOUSE	O08848 mus musculu
34	7	0.7	591	1	GV7_XENLA	P20398 xenopus lae
35	7	0.7	603	1	MTHR_SCHPO	Q10258 schizosacch
36	7	0.7	618	1	CAB2_METJA	Q58776 methanococc
37	7	0.7	624	1	NKX3_RAT	Q9epq0 rattus norv
38	7	0.7	644	1	NKX3_HUMAN	Q9hc58 homo sapien
39	7	0.7	645	1	NKX3_MOUSE	Q99pd7 mus musculu
40	7	0.7	677	1	CFPA_TREPA	Q56336 treponema p
41	7	0.7	717	1	NAH4_RAT	P26434 rattus norv
42	7	0.7	732	1	PRIA_ECOLI	P17888 escherichia
43	7	0.7	740	1	EF2_PYRAE	Q8zzc1 pyrobaculum
44	7	0.7	777	1	UNC8_CAEEL	Q21974 caenorhabdi
45	7	0.7	892	1	RA16_SCHPO	P36617 schizosacch
46	7	0.7	944	1	YMH6_YEAST	Q03631 saccharomyc
47	7	0.7	1024	1	POP1_HUMAN	Q99575 homo sapien
48	7	0.7	1040	1	MAN1_RAT	P21139 rattus norv
49	7	0.7	1254	1	POLS_EEVV8	P05674 venezuelan
50	7	0.7	1254	1	POLS_EEVVM	P36331 venezuelan
51	7	0.7	1254	1	POLS_EEVVT	P09592 venezuelan
52	7	0.7	1255	1	POLS_EEVV3	P36329 venezuelan
53	7	0.7	1255	1	POLS_EEVVP	P36332 venezuelan
54	7	0.7	2514	1	POLN_SINDO	P27283 sindbis vir
55	7	0.7	2549	1	FRAP_HUMAN	P42345 homo sapien
56	7	0.7	2554	1	7LES_DROME	P13368 drosophila
57	7	0.7	2594	1	7LES_DROVI	P20806 drosophila
58	7	0.7	2845	1	APC_MOUSE	Q61315 mus musculu
59	7	0.7	3829	1	SACS_HUMAN	Q9nzj4 homo sapien
60	7	0.7	3830	1	SACS_MOUSE	Q9jlc8 mus musculu
61	7	0.7	4644	1	DYHC_RAT	P38650 rattus norv
62	6	0.6	37	1	PSBL_ARATH	P29301 arabidopsis
63	6	0.6	37	1	TX21_SELHU	P82959 selenocosmi
64	6	0.6	37	1	TX22_SELHU	P82960 selenocosmi
65	6	0.6	38	1	PSBL_CHLVU	P56339 chlorella v
66	6	0.6	38	1	PSBL_MARPO	P12165 marchantia
67	6	0.6	43	1	PIV6_ADEB2	Q96627 bovine aden
68	6	0.6	48	1	ATP8_YARLI	Q36257 yarrowia li
69	6	0.6	54	1	ATP8_PARLI	P12697 paracentrot
70	6	0.6	67	1	Y737_ARCFU	O29521 archaeoglob
71	6	0.6	78	1	RUXF_SCHPO	O59734 schizosacch
72	6	0.6	79	1	NSGX_HUMAN	Q9uh64 homo sapien
73	6	0.6	81	1	YVKA_VACCC	P20569 vaccinia vi
74	6	0.6	83	1	Y67_BPT3	P20330 bacteriopha
75	6	0.6	86	1	Y425_TREPA	O83440 treponema p
76	6	0.6	88	1	FXY4_MOUSE	Q9d2w0 mus musculu
77	6	0.6	89	1	XHLA_BACSU	P39798 bacillus su
78	6	0.6	89	1	YAL1_BACLI	Q99164 bacillus li
79	6	0.6	95	1	ES6B_MYCTU	O05454 mycobacteri
80	6	0.6	97	1	XPA_CRIGR	Q64029 cricetulus
81	6	0.6	97	1	YAF1_BACLI	P37135 bacillus li
82	6	0.6	98	1	RK23_ASTLO	P34771 astasia lon
83	6	0.6	102	1	YNI6_YEAST	P48232 saccharomyc
84	6	0.6	105	1	GLN1_METTM	Q50786 methanobact
85	6	0.6	105	1	HIS3_LISIN	Q92e89 listeria in
86	6	0.6	105	1	THI1_SYNY3	P52232 synechocyst
87	6	0.6	108	1	Y793_SYNY3	Q55939 synechocyst
88	6	0.6	109	1	VIF_HV1SC	P05899 human immun
89	6	0.6	112	1	RLA1_DROME	P08570 drosophila
90	6	0.6	113	1	V195_FOWPV	Q9j538 fowlpox vir
91	6	0.6	115	1	HIS3_CLOAB	Q97kh7 clostridium
92	6	0.6	120	1	VP3_CAV26	P54095 chicken ane
93	6	0.6	121	1	SR14_ARATH	O04421 arabidopsis
94	6	0.6	121	1	VP3_CAV82	P54096 chicken ane
95	6	0.6	121	1	VP3_CAVC1	Q99152 chicken ane
96	6	0.6	121	1	VP3_CAVCI	P54094 chicken ane
97	6	0.6	124	1	VCL1_PEA	P02856 pisum sativ
98	6	0.6	127	1	ACPS_THETN	Q8r857 thermoanaer
99	6	0.6	128	1	CD59_AOTTR	P51447 aotus trivi
100	6	0.6	128	1	VL05_VACCV	P07615 vaccinia vi
101	6	0.6	128	1	VL05_VARV	P33043 variola vir
102	6	0.6	130	1	CLP1_DROME	P02839 drosophila
103	6	0.6	130	1	VMEM_NMV	P15097 narcissus m
104	6	0.6	131	1	IMEP_STRNI	P01077 streptomyce
105	6	0.6	133	1	ACPS_CLOPE	Q8xnp1 clostridium
106	6	0.6	133	1	RL15_HELPJ	Q9zjs8 helicobacte

107	6	0.6	134	1	YBIM_ECOLI	P75781 escherichia
108	6	0.6	135	1	RL15_HELPY	P56040 helicobacte
109	6	0.6	135	1	SPEH_AQUAE	O66615 aquifex aeo
110	6	0.6	139	1	ACPS_BRAJA	O69159 bradyrhizob
111	6	0.6	140	1	GTH2_ANGAN	P27767 anguilla an
112	6	0.6	141	1	LYSB_BPP2	P51770 bacteriopha
113	6	0.6	142	1	YE72_SCHPO	O14155 schizosacch
114	6	0.6	142	1	YNK5_YEAST	P50943 saccharomyc
115	6	0.6	145	1	ANG3_MOUSE	P97802 mus musculu
116	6	0.6	146	1	HBE_DIDMA	P11025 didelphis m
117	6	0.6	148	1	YMS3_CAEEL	P34499 caenorhabdi
118	6	0.6	148	1	YMS4_CAEEL	P34500 caenorhabdi
119	6	0.6	150	1	ARG2_CLOPE	P58685 clostridium
120	6	0.6	150	1	YFHP_HAEIN	P44675 haemophilus
121	6	0.6	152	1	SP1A_STRPU	P04109 strongyloce
122	6	0.6	154	1	PFDA_AERPE	Q9yd28 aeropyrum p
123	6	0.6	157	1	VE6_HPV36	P50810 human papil
124	6	0.6	159	1	R11G_ARATH	O65569 arabidopsis
125	6	0.6	160	1	YGE7_PSEAE	Q9htu7 pseudomonas
126	6	0.6	162	1	G16O_BACSU	P80872 bacillus su
127	6	0.6	162	1	SIEB_ECOLI	P38392 escherichia
128	6	0.6	163	1	DNE1_CHLRE	P05725 chlamydomon
129	6	0.6	164	1	YOR2_AZOVI	O69219 azotobacter
130	6	0.6	165	1	TPX_HAEIN	Q57549 haemophilus
131	6	0.6	166	1	PSAL_SYNP7	P95822 synechococc
132	6	0.6	170	1	RL17_MYCLE	Q9x797 mycobacteri
133	6	0.6	171	1	IPP1_HUMAN	Q13522 homo sapien
134	6	0.6	173	1	AROK_BUCAI	P57605 buchnera ap
135	6	0.6	175	1	IPYR_SALTY	Q8xgi0 salmonella
136	6	0.6	176	1	IPYR_VIBCH	Q9kp34 vibrio chol
137	6	0.6	177	1	AAC1_PSEAE	P23181 pseudomonas
138	6	0.6	177	1	ATPD_BUCAP	O51875 buchnera ap
139	6	0.6	181	1	FANC_ECOLI	P18103 escherichia
140	6	0.6	183	1	KITH_FOWPV	P10052 fowlpox vir
141	6	0.6	187	1	YKT9_YEAST	P36045 saccharomyc
142	6	0.6	189	1	OXYR_MYCXE	O87883 mycobacteri
143	6	0.6	190	1	SOMA_PAROL	P09537 paralichthy
144	6	0.6	195	1	PAPH_ECOLI	P07111 escherichia
145	6	0.6	195	1	PRSH_ECOLI	P42185 escherichia
146	6	0.6	196	1	HIR5_HUMAN	Q9ums0 homo sapien
147	6	0.6	197	1	CCP1_CRIGR	O35847 cricetulus
148	6	0.6	197	1	MOBA_METTH	O26246 methanobact
149	6	0.6	198	1	ATPF_MYCGA	P33256 mycoplasma
150	6	0.6	199	1	HIR5_MOUSE	Q9qz23 mus musculu

ALIGNMENTS

```
RESULT 1
KIME_ARCFU
ID   KIME_ARCFU     STANDARD;      PRT;   284 AA.
AC   O27995;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mevalonate kinase (EC 2.7.1.36) (MK).
GN   MVK OR AF2289.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- CATALYTIC ACTIVITY: ATP + (R)-mevalonate = ADP + (R)-5-
CC       phosphomevalonate.
CC   -!- COFACTOR: MAGNESIUM (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE GHMP KINASE FAMILY. MEVALONATE KINASE
CC       SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000946; AAB88965.1; -.
DR   TIGR; AF2289; -.
DR   InterPro; IPR001745; GHMPknse_ATP.
DR   InterPro; IPR001459; Mev_gal_kin.
DR   Pfam; PF00288; GHMP_kinases; 1.
DR   PRINTS; PR00959; MEVGALKINASE.
DR   TIGRFAMs; TIGR00549; mevalon_kin; 1.
DR   PROSITE; PS00627; GHMP_KINASES_ATP; 1.
KW   Transferase; Kinase; ATP-binding; Magnesium; Complete proteome.
FT   NP_BIND      86     96       ATP (POTENTIAL).
SQ   SEQUENCE   284 AA;  30877 MW;  F102C1C71772CA7F CRC64;

  Query Match             0.8%;  Score 8;  DB 1;  Length 284;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     593 LEASDVGS 600
           ||||||||
Db     201 LEASDVGS 208


RESULT 2
PERA_ARMRU
ID   PERA_ARMRU     STANDARD;      PRT;   353 AA.
AC   P00433;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Peroxidase C1A precursor (EC 1.11.1.7).
GN   PRXC1A OR HPRC1.
OS   Armoracia rusticana (Horseradish) (Armoracia laphatifolia).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Armoracia.
OX   NCBI_TaxID=3704;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88225087; PubMed=3371352;
RA   Fujiyama K., Takemura H., Shibayama S., Kobayashi K., Choi J.K.,
RA   Shinmyo A., Takano M., Yamada Y., Okada H.;
RT   "Structure of the horseradish peroxidase isozyme C genes.";
RL   Eur. J. Biochem. 173:681-687(1988).
RN   [2]
RP   SEQUENCE OF 31-338.
RX   MEDLINE=77068850; PubMed=1001465;
RA   Welinder K.G.;
RT   "Covalent structure of the glycoprotein horseradish peroxidase (EC
RT   1.11.1.7).";
RL   FEBS Lett. 72:19-23(1976).
RN   [3]
```

```
RP   X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
RX   MEDLINE=98069652; PubMed=9406554;
RA   Gajhede M., Schuller D.J., Henriksen A., Smith A.T., Poulos T.L.;
RT   "Crystal structure of horseradish peroxidase C at 2.15-A resolution.";
RL   Nat. Struct. Biol. 4:1032-1038(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX   MEDLINE=98272679; PubMed=9609699;
RA   Henriksen A., Schuller D.J., Meno K., Welinder K.G., Smith A.T.,
RA   Gajhede M.;
RT   "Structural interactions between horseradish peroxidase C and the
RT   substrate benzhydroxamic acid determined by X-ray crystallography.";
RL   Biochemistry 37:8054-8060(1998).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RA   Meno K., White C.G., Smith A.T., Gajhede M.;
RL   Submitted (DEC-1998) to the PDB data bank.
CC   -!- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
CC       BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD
CC       WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
CC       DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
CC   -!- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC   -!- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 2 calcium ions
CC       per subunit.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. PLANT PEROXIDASE
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37156; AAA33377.1; ALT_SEQ.
DR   PIR; A00502; OPRHC.
DR   PIR; S00625; S00625.
DR   PDB; 1ATJ; 04-FEB-98.
DR   PDB; 2ATJ; 28-JAN-98.
DR   PDB; 3ATJ; 23-DEC-98.
DR   GlycoSuiteDB; P00433; -.
DR   InterPro; IPR002016; Peroxidase.
DR   Pfam; PF00141; peroxidase; 1.
DR   PRINTS; PR00458; PEROXIDASE.
DR   PROSITE; PS00435; PEROXIDASE_1; 1.
DR   PROSITE; PS00436; PEROXIDASE_2; 1.
KW   Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
KW   Multigene family; Signal; 3D-structure.
FT   SIGNAL       1     30
FT   CHAIN       31    338       PEROXIDASE C1A.
FT   PROPEP     339    353
FT   CA_BIND      73     73       1.
FT   CA_BIND      76     76       1 (VIA CARBONYL OXYGEN).
FT   CA_BIND      78     78       1 (VIA CARBONYL OXYGEN).
FT   CA_BIND      80     80       1.
FT   CA_BIND      82     82       1.
FT   CA_BIND      94     94       1.
FT   CA_BIND     201    201       2.
FT   CA_BIND     236    252       2.
FT   CA_BIND     239    254       2.
FT   CA_BIND     242    257       2.
FT   CA_BIND     244    259       2.
FT   ACT_SITE     68     68
FT   ACT_SITE     72     72       DISTAL HISTIDINE.
FT   METAL       200    200       IRON (PROTOHEME IX AXIAL LIGAND).
FT   MOD_RES      31     31       PYRROLIDONE CARBOXYLIC ACID.
FT   DISULFID     41    121
FT   DISULFID     74     79
FT   DISULFID    127    331
FT   DISULFID    207    239
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    188    188       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    216    216       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .).
SQ   SEQUENCE   353 AA;  38825 MW;  AC916C03C4A24D27 CRC64;

  Query Match             0.8%;  Score 8;  DB 1;  Length 353;
  Best Local Similarity   100.0%;  Pred. No. 3.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     680 VTLAGGPS 687
           ||||||||
Db     139 VTLAGGPS 146


RESULT 3
PERC_ARATH
ID   PERC_ARATH     STANDARD;      PRT;   354 AA.
AC   P24101;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neutral peroxidase C precursor (EC 1.11.1.7).
GN   PRXCA OR AT3G49110 OR T2J13.50.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91200671; PubMed=2016063;
RA   Intapruk C., Higashimura N., Yamamoto K., Okada N., Shinmyo A.,
RA   Takano M.;
RT   "Nucleotide sequences of two genomic DNAs encoding peroxidase of
RT   Arabidopsis thaliana.";
RL   Gene 98:237-241(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
```

```
CC   -!- FUNCTION: REMOVAL OF H(2)O(2), OXIDATION OF TOXIC REDUCTANTS,
CC       BIOSYNTHESIS AND DEGRADATION OF LIGNIN, DEFENSE RESPONSE TOWARD
CC       WOUNDING AND METABOLISM OF AUXIN. THESE FUNCTIONS MIGHT BE
CC       DEPENDENT ON EACH ISOZYME/ISOFORM IN EACH PLANT TISSUE.
CC   -!- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC   -!- COFACTOR: Binds 1 protoheme IX and 1 iron(III) ion.
CC   -!- TISSUE SPECIFICITY: ROOTS.
CC   -!- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. PLANT PEROXIDASE
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58380; AAA32849.1; -.
DR   EMBL; AL132967; CAB61999.1; -.
DR   PIR; JU0457; JU0457.
DR   HSSP; P00433; 2ATJ.
DR   InterPro; IPR002016; Peroxidase.
DR   Pfam; PF00141; peroxidase; 1.
DR   PRINTS; PR00458; PEROXIDASE.
DR   PROSITE; PS00435; PEROXIDASE_1; 1.
DR   PROSITE; PS00436; PEROXIDASE_2; 1.
KW   Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
KW   Multigene family; Signal.
FT   SIGNAL       1     21
FT   CHAIN       22    354       NEUTRAL PEROXIDASE C.
FT   ACT_SITE    69     69       BY SIMILARITY.
FT   ACT_SITE    73     73       DISTAL HISTIDINE (BY SIMILARITY).
FT   METAL      201    201       IRON (PROTOHEME IX AXIAL LIGAND).
FT   DISULFID    42    122       BY SIMILARITY.
FT   DISULFID    75     80       BY SIMILARITY.
FT   DISULFID   128    332       BY SIMILARITY.
FT   DISULFID   208    240       BY SIMILARITY.
FT   CARBOHYD    88     88       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   217    217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   245    245       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   286    286       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   354 AA;  38941 MW;  6D2EE3D536111724 CRC64;

  Query Match             0.8%;  Score 8;  DB 1;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 3.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     680 VTLAGGPS 687
           ||||||||
Db     140 VTLAGGPS 147


RESULT 4
TOLB_RICPR
ID   TOLB_RICPR     STANDARD;      PRT;   443 AA.
AC   Q9ZDM5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TolB protein precursor.
GN   TOLB OR RP302.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE TOLB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235271; CAA14763.1; -.
KW   Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT   SIGNAL       1     18       POTENTIAL.
FT   CHAIN       19    443       TOLB PROTEIN.
SQ   SEQUENCE   443 AA;  49206 MW;  BE5BC2282A0F5593 CRC64;

  Query Match             0.8%;  Score 8;  DB 1;  Length 443;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     161 PRGDYIAF 168
           ||||||||
Db     340 PRGDYIAF 347


RESULT 5
DYHC_MOUSE
ID   DYHC_MOUSE     STANDARD;      PRT;  4644 AA.
AC   Q9JHU4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain).
GN   DNCHC1 OR DNCH1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB;
RA   Sasaki S., Shionoya A., Hirotsune S.;
RT   "Complete cDNA sequence of murine cytoplasmic dynein heavy chain.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC       MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC       ORGANELLES ALONG MICROTUBULES.
CC   -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC       INTERMEDIATE AND LIGHT CHAINS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY004877; AAF91078.1; -.
DR   MGD; MGI:103147; Dnchc1.
DR   InterPro; IPR004273; Dynein_heavy.
DR   Pfam; PF03028; Dynein_heavy; 1.
KW   Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT   DOMAIN       48     69       COILED COIL (POTENTIAL).
```

```
FT   DOMAIN      179    200       COILED COIL (POTENTIAL).
FT   DOMAIN      453    476       COILED COIL (POTENTIAL).
FT   DOMAIN      541    564       COILED COIL (POTENTIAL).
FT   DOMAIN     1169   1201       COILED COIL (POTENTIAL).
FT   DOMAIN     1229   1250       COILED COIL (POTENTIAL).
FT   DOMAIN     1355   1371       COILED COIL (POTENTIAL).
FT   DOMAIN     2012   2040       MICROTUBULE-BINDING (POTENTIAL).
FT   DOMAIN     3187   3273       COILED COIL (POTENTIAL).
FT   DOMAIN     3394   3498       COILED COIL (POTENTIAL).
FT   DOMAIN     3735   3798       COILED COIL (POTENTIAL).
FT   NP_BIND    1904   1911       ATP (POTENTIAL).
FT   NP_BIND    2222   2229       ATP (POTENTIAL).
FT   NP_BIND    2593   2600       ATP (POTENTIAL).
FT   NP_BIND    2935   2942       ATP (POTENTIAL).
SQ   SEQUENCE   4644 AA;  532021 MW;  FE5B4E15DD479E1B CRC64;

  Query Match             0.8%;  Score 8;  DB 1;  Length 4644;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 SARVRGRT 18
           ||||||||
Db     720 SARVRGRT 727


RESULT 6
NER_BPD10
ID   NER_BPD10      STANDARD;      PRT;    73 AA.
AC   P06903;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-binding protein Ner.
GN   NER.
OS   Bacteriophage D108.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC   Mu-like viruses.
OX   NCBI_TaxID=10671;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86055744; PubMed=2998774;
RA   Tolias P.P., Dubow M.S.;
RT   "The cloning and characterization of the bacteriophage D108
RT   regulatory DNA-binding protein ner.";
RL   EMBO J. 4:3031-3037(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86232621; PubMed=3012481;
RA   Mizuuchi M., Weisberg R.A., Mizuuchi K.;
RT   "DNA sequence of the control region of phage D108: the N-terminal
RT   amino acid sequences of repressor and transposase are similar both in
RT   phage D108 and in its relative, phage Mu.";
RL   Nucleic Acids Res. 14:3813-3825(1986).
CC   -!- FUNCTION: BINDS TO THE REGION RIGHT AFTER THE START OF
CC       TRANSCRIPTION OF THE PROMOTER REGION OF THE EARLY GENES.
CC   -!- SIMILARITY: BELONGS TO THE NER FAMILY OF TRANSCRIPTIONAL
CC       REGULATORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M26291; AAA32206.1; -.
DR   EMBL; X03847; CAA27475.1; -.
DR   PIR; A24680; DNBPD8.
DR   PIR; S07931; S07931.
DR   HSSP; P06020; 1NEQ.
KW   Transcription regulation; DNA-binding.
FT   DNA_BIND     28     47       H-T-H MOTIF (PROBABLE).
SQ   SEQUENCE   73 AA;  8534 MW;  898DFD38E1409025 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     982 LGRSNHL 988
           |||||||
Db      31 LGRSNHL 37


RESULT 7
LSM6_SCHPO
ID   LSM6_SCHPO     STANDARD;      PRT;    75 AA.
AC   Q9UUI1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable U6 snRNA-associated Sm-like protein LSm6.
GN   SPAC2F3.17C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z99165; CAB54975.1; -.
DR   InterPro; IPR001163; snRNP_Sm.
DR   Pfam; PF01423; Sm; 1.
KW   Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
KW   RNA-binding.
SQ   SEQUENCE   75 AA;  8336 MW;  20B1B1F2E380BCCF CRC64;
```

```
  Query Match             0.7%;  Score 7;  DB 1;  Length 75;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     237 RGNNVLY 243
           |||||||
Db      63 RGNNVLY 69


RESULT 8
LSM6_HUMAN
ID   LSM6_HUMAN      STANDARD;      PRT;     80 AA.
AC   Q9Y4Y8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   U6 snRNA-associated Sm-like protein LSm6.
GN   LSM6.
OS   Homo sapiens (Human), and
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606, 10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human;
RX   MEDLINE=99298196; PubMed=10369684;
RA   Salgado-Garrido J., Bragado-Nilsson E., Kandels-Lewis S., Seraphin B.;
RT   "Sm and Sm-like proteins assemble in two related complexes of deep
RT   evolutionary origin.";
RL   EMBO J. 18:3451-3462(1999).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   SPECIES=Human;
RX   MEDLINE=99452783; PubMed=10523320;
RA   Achsel T., Brahms H., Kastner B., Bachi A., Wilm M., Luehrmann R.;
RT   "A doughnut-shaped heteromer of human Sm-like proteins binds to the
RT   3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in
RT   vitro.";
RL   EMBO J. 18:5789-5802(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Mouse; STRAIN=C57BL/6J;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: BINDS SPECIFICALLY TO THE 3'-TERMINAL U-TRACT OF U6
CC       SNRNA.
CC   -!- SUBUNIT: LSM SUBUNITS FORM A HETEROMER WITH A DOUGNUT SHAPE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ238098; CAB45869.1; -.
DR   EMBL; AF182292; AAD56230.1; -.
DR   EMBL; AK019126; BAB31555.1; -.
DR   InterPro; IPR001163; snRNP_Sm.
DR   Pfam; PF01423; Sm; 1.
KW   Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
KW   RNA-binding.
SQ   SEQUENCE   80 AA;  9128 MW;  21167891FDE804F1 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     237 RGNNVLY 243
           |||||||
Db      66 RGNNVLY 72


RESULT 9
URE3_ECOLI
ID   URE3_ECOLI      STANDARD;      PRT;    100 AA.
AC   Q03282;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN   UREA.
OS   Escherichia coli.
OG   Plasmid.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93194816; PubMed=8449894;
RA   D'Orazio S.E., Collins C.M.;
RT   "Characterization of a plasmid-encoded urease gene cluster found in
RT   members of the family Enterobacteriaceae.";
RL   J. Bacteriol. 175:1860-1864(1993).
CC   -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC   -!- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L03307; AAA24745.1; -.
DR   PIR; B47090; B47090.
DR   HSSP; P18316; 1FWB.
DR   InterPro; IPR002026; Urease_gamma.
DR   Pfam; PF00547; urease_gamma; 1.
DR   ProDom; PD002319; Urease_gamma; 1.
DR   TIGRFAMs; TIGR00193; urease_gam; 1.
KW   Hydrolase; Plasmid.
SQ   SEQUENCE   100 AA;  10983 MW;  FBFD74F4C86B0805 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     960 SCAIMEG 966
           |||||||
Db      40 SCAIMEG 46
```

```
RESULT 10
URE3_PROMI
ID   URE3_PROMI     STANDARD;      PRT;   100 AA.
AC   P17088;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN   UREA.
OS   Proteus mirabilis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=584;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HI4320;
RX   MEDLINE=90078080; PubMed=2687233;
RA   Jones B.D., Mobley H.L.T.;
RT   "Proteus mirabilis urease: nucleotide sequence determination and
RT   comparison with jack bean urease.";
RL   J. Bacteriol. 171:6414-6422(1989).
CC   -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC   -!- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31834; AAA25667.1; -.
DR   PIR; B43719; B43719.
DR   HSSP; P18316; 1FWB.
DR   InterPro; IPR002026; Urease_gamma.
DR   Pfam; PF00547; urease_gamma; 1.
DR   ProDom; PD002319; Urease_gamma; 1.
DR   TIGRFAMs; TIGR00193; urease_gam; 1.
KW   Hydrolase.
SQ   SEQUENCE   100 AA;  10998 MW;  79E4A2182C1EF903 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    960 SCAIMEG 966
          |||||||
Db     40 SCAIMEG 46


RESULT 11
URE3_PROVU
ID   URE3_PROVU     STANDARD;      PRT;   100 AA.
AC   P16124;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN   UREA.
OS   Proteus vulgaris.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=585;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 30118;
RX   MEDLINE=90215174; PubMed=2182382;
RA   Moersdorf G., Kaltwasser H.;
RT   "Cloning of the genes encoding urease from Proteus vulgaris and
RT   sequencing of the structural genes.";
RL   FEMS Microbiol. Lett. 54:67-73(1990).
CC   -!- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC   -!- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51816; CAA36113.1; -.
DR   PIR; S08478; S08478.
DR   HSSP; P18316; 1FWB.
DR   InterPro; IPR002026; Urease_gamma.
DR   Pfam; PF00547; urease_gamma; 1.
DR   ProDom; PD002319; Urease_gamma; 1.
DR   TIGRFAMs; TIGR00193; urease_gam; 1.
KW   Hydrolase.
SQ   SEQUENCE   100 AA;  10955 MW;  D9F413A9DA2C84B5 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    960 SCAIMEG 966
          |||||||
Db     40 SCAIMEG 46


RESULT 12
DEF_RICCN
ID   DEF_RICCN      STANDARD;      PRT;   175 AA.
AC   Q92IZ1;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN   DEF OR RC0278.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
CC   -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC       newly synthesized proteins. Requires at least a dipeptide for an
CC       efficient rate of reaction. N-terminal L-methionine is a
CC       prerequisite for activity but the enzyme has broad specificity at
CC       other positions (By similarity).
CC   -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC       methionyl peptide.
CC   -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; AE008593; AAL02816.1; -.
DR   InterPro; IPR000181; Pep_deformylase.
DR   Pfam; PF01327; Pep_deformylase; 1.
DR   ProDom; PD003844; Pep_deformylase; 1.
DR   TIGRFAMs; TIGR00079; pept_deformyl; 1.
KW   Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT   METAL        99     99       IRON (BY SIMILARITY).
FT   METAL       141    141       IRON (BY SIMILARITY).
FT   ACT_SITE    142    142       BY SIMILARITY.
FT   METAL       145    145       IRON (BY SIMILARITY).
SQ   SEQUENCE   175 AA;  20279 MW;  93B050C6003C5A6D CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     892 ISLPEQR 898
           |||||||
Db     100 ISLPEQR 106


RESULT 13
Y773_METJA
ID   Y773_METJA     STANDARD;      PRT;   175 AA.
AC   Q58183;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0773 precursor.
GN   MJ0773.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67522; AAB98777.1; -.
DR   TIGR; MJ0773; -.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL       1     33       POTENTIAL.
FT   CHAIN       34    175       HYPOTHETICAL PROTEIN MJ0773.
SQ   SEQUENCE   175 AA;  20246 MW;  63EE09A737EE4F11 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 175;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     687 SFTSKGL 693
           |||||||
Db      68 SFTSKGL 74


RESULT 14
YCF4_ODOSI
ID   YCF4_ODOSI     STANDARD;      PRT;   181 AA.
AC   P49526;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Photosystem I assembly protein ycf4.
GN   YCF4.
OS   Odontella sinensis (Marine centric diatom).
OG   Chloroplast.
OC   Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC   Biddulphiophycidae; Eupodiscales; Eupodiscaceae; Odontella.
OX   NCBI_TaxID=2839;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kowallik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT   "The chloroplast genome of a chlorophyll a+c-containing alga,
RT   Odontella sinensis.";
RL   Plant Mol. Biol. Rep. 13:336-342(1995).
CC   -!- FUNCTION: Seems to be required for the assembly of the photosystem
CC       I complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Thylakoid membrane-associated (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE YCF4 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z67753; CAA91709.1; -.
DR   InterPro; IPR003359; PSI_Ycf4.
DR   Pfam; PF02392; Ycf4; 1.
KW   Photosynthesis; Thylakoid; Transmembrane; Chloroplast.
FT   TRANSMEM     19     41       POTENTIAL.
FT   TRANSMEM     61     83       POTENTIAL.
SQ   SEQUENCE   181 AA;  20547 MW;  663832A7DB554EFA CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 181;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     887 LCSGGIS 893
           |||||||
Db      25 LCSGGIS 31


RESULT 15
PMG2_ECOLI
ID   PMG2_ECOLI     STANDARD;      PRT;   215 AA.
AC   P36942;
DT   01-JUN-1994 (Rel. 29, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable phosphoglycerate mutase 2 (EC 5.4.2.1) (Phosphoglyceromutase
DE   2) (PGAM 2) (BPG-dependent PGAM 2).
GN   GPMB OR B4395 OR Z5997 OR ECS5353.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RX   MEDLINE=93194823; PubMed=8449900;
RA   Skarstad K., Thoeny B., Hwang D.S., Kornberg A.;
RT   "A novel binding protein of the origin of the Escherichia coli
```

```
RT   chromosome.";
RL   J. Biol. Chem. 268:5365-5370(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA   Blattner F.R.;
RT   "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT   region from 92.8 through 100 minutes.";
RL   Nucleic Acids Res. 23:2105-2119(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate + 2,3-diphosphoglycerate
CC       = 3-phospho-D-glycerate + 2,3-diphosphoglycerate.
CC   -!- PATHWAY: Glycolysis.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE MUTASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M97495; -; NOT_ANNOTATED_CDS.
DR   EMBL; U14003; AAA97291.1; -.
DR   EMBL; AE000509; AAC77348.1; -.
DR   EMBL; AE005670; AAG59575.1; -.
DR   EMBL; AP002569; BAB38776.1; -.
DR   EcoGene; EG12164; gpmB.
DR   InterPro; IPR001345; PG/BPGM_mutase.
DR   Pfam; PF00300; PGAM; 1.
DR   PROSITE; PS00175; PG_MUTASE; 1.
KW   Isomerase; Glycolysis; Multigene family; Complete proteome.
FT   ACT_SITE      9      9       FORMS THE PHOSPHOHISTIDINE INTERMEDIATE
FT                                (BY SIMILARITY).
FT   ACT_SITE     58     58       REQUIRED FOR BINDING CARBOXYL GROUP OF
FT                                PHOSPHOGLYCERATES (BY SIMILARITY).
FT   ACT_SITE    150    150       BY SIMILARITY.
FT   CONFLICT     35     35       Q -> H (IN REF. 3).
FT   CONFLICT    148    148       V -> L (IN REF. 1).
SQ   SEQUENCE   215 AA;  24065 MW;  3653DA0548B9E009 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     722 RIPEGES 728
           |||||||
Db     115 RIPEGES 121
```

```
RESULT 16
DHSB_RECAM
ID   DHSB_RECAM     STANDARD;      PRT;   239 AA.
AC   P80480;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Succinate dehydrogenase [ubiquinone] iron-sulfur protein (EC 1.3.5.1)
DE   (IP).
GN   SDH2 OR SDHB.
OS   Reclinomonas americana.
OG   Mitochondrion.
OC   Eukaryota; core jakobids; Reclinomonas.
OX   NCBI_TaxID=48483;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 50394;
RX   MEDLINE=96197311; PubMed=8637872;
RA   Burger G., Lang B.F., Reith M., Gray M.W.;
RT   "Genes encoding the same three subunits of respiratory complex II are
RT   present in the mitochondrial DNA of two phylogenetically distant
RT   eukaryotes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:2328-2332(1996).
CC   -!- CATALYTIC ACTIVITY: Succinate + ubiquinone = fumarate + ubiquinol.
CC   -!- COFACTOR: BINDS THREE DIFFERENT IRON-SULFUR CLUSTERS: A 2FE-2S,
CC       A 3FE-4S AND A 4FE-4S (BY SIMILARITY).
CC   -!- PATHWAY: Tricarboxylic acid cycle.
CC   -!- SUBUNIT: COMPOSED OF A 27 kDa IRON PROTEIN (IP) AND A 70 kDa
CC       FLAVOPROTEIN (FP).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF007261; AAD11913.1; -.
DR   HSSP; P00364; 1FUM.
DR   InterPro; IPR000564; 2Fe2S_ferredoxin.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR004489; DhsB.
DR   InterPro; IPR001041; Ferredoxin.
DR   Pfam; PF00111; fer2; 1.
DR   TIGRFAMs; TIGR00384; dhsB; 1.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW   Tricarboxylic acid cycle; Iron-sulfur; Oxidoreductase; Mitochondrion;
KW   Electron transport; 3Fe-4S; 4Fe-4S.
FT   METAL        61     61       IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT   METAL        66     66       IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT   METAL        69     69       IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT   METAL        81     81       IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT   METAL       151    151       IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT   METAL       154    154       IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT   METAL       157    157       IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT   METAL       161    161       IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
FT   METAL       208    208       IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
FT   METAL       214    214       IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
FT   METAL       218    218       IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   239 AA;  27828 MW;  781C6FF0ED55375B CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 239;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     220 WMKTTEK 226
           |||||||
Db     123 WMKTTEK 129
```

```
RESULT 17
IF6_SCHPO
ID   IF6_SCHPO      STANDARD;      PRT;   244 AA.
AC   O94476;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Eukaryotic translation initiation factor 6 (eIF-6).
GN   TIF6 OR SPCC1919.09.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: BINDS TO THE 60S RIBOSOMAL SUBUNIT AND PREVENTS ITS
CC       ASSOCIATION WITH THE 40S RIBOSOMAL SUBUNIT TO FORM THE 80S
CC       INITIATION COMPLEX (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE EIF-6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL035075; CAA22640.1; -.
DR   HSSP; Q12522; 1G62.
DR   InterPro; IPR002769; eIF6.
DR   Pfam; PF01912; eIF6; 1.
DR   ProDom; PD006880; eIF6; 1.
KW   Initiation factor; Protein biosynthesis.
SQ   SEQUENCE   244 AA;  26232 MW;  388F52C670F25A3C CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 244;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      81 SLPDPVK 87
           |||||||
Db      87 SLPDPVK 93


RESULT 18
TRY7_ANOGA
ID   TRY7_ANOGA     STANDARD;      PRT;   267 AA.
AC   P35041;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Trypsin 7 precursor (EC 3.4.21.4).
GN   TRYP7.
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Culicoidea; Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Suakoko; TISSUE=Midgut;
RX   MEDLINE=93327779; PubMed=8335004;
RA   Mueller H.M., Crampton J.M., della Torre A., Sinden R., Crisanti A.;
RT   "Members of a trypsin gene family in Anopheles gambiae are induced in
RT   the gut by blood meal.";
RL   EMBO J. 12:2891-2900(1993).
CC   -!- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD
CC       MEAL.
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: MIDGUT.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z22930; CAA80516.1; -.
DR   PIR; S40006; S40006.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.112; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Signal; Zymogen; Digestion;
KW   Multigene family.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19     41       ACTIVATION PEPTIDE.
FT   CHAIN        42    267       TRYPSIN 7.
FT   ACT_SITE     82     82       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    127    127       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    222    222       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     67     83       BY SIMILARITY.
FT   DISULFID    192    207       BY SIMILARITY.
FT   DISULFID    218    242       BY SIMILARITY.
FT   SITE        216    216       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE   267 AA;  28456 MW;  11015871257DB678 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     920 AILLTVL 926
           |||||||
Db       6 AILLTVL 12
```

```
RESULT 19
YC15_PYRHO
ID   YC15_PYRHO     STANDARD;      PRT;   292 AA.
AC   O58968;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical ABC transporter permease protein PH1215.
GN   PH1215 OR PHBK039.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM PH1214/15/16. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION
CC       OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM PERMEASE FAMILY. MALFG SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000005; BAA30315.1; -.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF00528; BPD_transp; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
KW   Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT   TRANSMEM     10     30       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM    106    126       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
FT   TRANSMEM    215    235       POTENTIAL.
FT   TRANSMEM    261    281       POTENTIAL.
SQ   SEQUENCE   292 AA;  33407 MW;  5B968812C0B89F68 CRC64;

  Query Match              0.7%;  Score 7;  DB 1;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     826 VPGSLLL 832
           |||||||
Db      79 VPGSLLL 85


RESULT 20
RRP4_HUMAN
ID   RRP4_HUMAN     STANDARD;      PRT;   293 AA.
AC   Q13868; Q9NUY4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Exosome complex exonuclease RRP4 (EC 3.1.13.-) (Ribosomal RNA
DE   processing protein 4).
GN   RRP4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung carcinoma;
RA   Chissoe S.L.;
RL   Thesis (1994), University of Oklahoma, U.S.A.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta, and Mammary gland;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND REVISIONS.
RC   TISSUE=Cervical adenocarcinoma;
RX   MEDLINE=98050921; PubMed=9390555;
RA   Mitchell P., Petfalski E., Shevchenko A., Mann M., Tollervey D.;
RT   "The exosome: a conserved eukaryotic RNA processing complex containing
RT   multiple 3'-->5' exoribonucleases.";
RL   Cell 91:457-466(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=96178774; PubMed=8600032;
RA   Mitchell P., Petfalski E., Tollervey D.;
RT   "The 3' end of yeast 5.8S rRNA is generated by an exonuclease
RT   processing mechanism.";
RL   Genes Dev. 10:502-513(1996).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=99396719; PubMed=10465791;
RA   Allmang C., Petfalski E., Podtelejnikov A., Mann M., Tollervey D.,
RA   Mitchell P.;
RT   "The yeast exosome and human PM-Scl are related complexes of 3'-->5'
RT   exonucleases.";
RL   Genes Dev. 13:2148-2158(1999).
CC   -!- FUNCTION: BELONGS TO THE EXOSOME, A RNA PROCESSING COMPLEX, WHICH
CC       IS AT LEAST INVOLVED IN THE 3' PROCESSING OF THE 7S PRE-RRNA TO
CC       THE MATURE 5.8S RRNA. EXHIBITS A 3'-5' EXORIBONUCLEASE ACTIVITY
CC       (BY SIMILARITY).
CC   -!- SUBUNIT: COMPONENT OF THE EXOSOME MULTIENZYME RIBONUCLEASE COMPLEX
CC       COMPOSED OF AT LEAST 11 PROTEINS: RRP4, RRP40, RRP41/SKI6, RRP42,
CC       RRP43, RRP44/DIS3, PM/SCL-75, RRP46, CSL4 AND PM/SCL-100 (ONLY IN
CC       THE NUCLEAR COMPLEX). ALSO ASSOCIATED WITH THE GTPASE RAN (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 S1 MOTIF DOMAIN.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO THE
CC       SKIPPING OF AN EXON.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U07561; AAB60392.1; ALT_SEQ.
DR   EMBL; AK001916; BAA91977.1; -.
DR   EMBL; AK022460; BAB14043.1; -.
DR   EMBL; BC000747; AAH00747.1; -.
```

```
DR   MIM; 602238; -.
DR   InterPro; IPR003029; S1.
DR   SMART; SM00316; S1; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; rRNA processing;
KW   Nuclear protein; RNA-binding.
FT   DOMAIN       79    159       S1 MOTIF.
SQ   SEQUENCE   293 AA;  32789 MW;  882033F50791643F CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     759 PVSLADR 765
           |||||||
Db     225 PVSLADR 231


RESULT 21
DPPC_ECOLI
ID   DPPC_ECOLI     STANDARD;      PRT;   300 AA.
AC   P37315;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dipeptide transport system permease protein dppC.
GN   DPPC OR B3542 OR Z4959 OR ECS4422.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MM500;
RX   MEDLINE=95231288; PubMed=7536291;
RA   Abouhamad W.N., Manson M.D.;
RT   "The dipeptide permease of Escherichia coli closely resembles other
RT   bacterial transport systems and shows growth-phase-dependent
RT   expression.";
RL   Mol. Microbiol. 14:1077-1092(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94316500; PubMed=8041620;
RA   Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   Nucleic Acids Res. 22:2576-2586(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR DIPEPTIDES; PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF
CC       THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM PERMEASE FAMILY. OPPBC SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08399; AAA23704.1; -.
DR   EMBL; U00039; AAB18520.1; -.
DR   EMBL; AE000431; AAC76567.1; -.
DR   EMBL; AE005580; AAG58686.1; -.
DR   EMBL; AP002565; BAB37845.1; -.
DR   EcoGene; EG12626; dppC.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF00528; BPD_transp; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; FALSE_NEG.
KW   Transport; Peptide transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     33     53       POTENTIAL.
FT   TRANSMEM    102    122       POTENTIAL.
FT   TRANSMEM    137    157       POTENTIAL.
FT   TRANSMEM    207    227       POTENTIAL.
FT   TRANSMEM    231    251       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
SQ   SEQUENCE   300 AA;  32308 MW;  C4DD7BD82286EE62 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     920 AILLTVL 926
           |||||||
Db     275 AILLTVL 281


RESULT 22
STBA_ECOLI
ID   STBA_ECOLI     STANDARD;      PRT;   320 AA.
AC   P11904; P10028;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Protein StbA (ParA locus 36 kDa protein).
GN   STBA.
OS   Escherichia coli.
OG   Plasmid IncFII NR1, and Plasmid IncFII R1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KP245; PLASMID=IncFII NR1;
RX   MEDLINE=89011976; PubMed=3172224;
RA   Tabuchi A., Min Y.-N., Kim C.K., Fan Y.-L., Womble D.D., Rownd R.H.;
RT   "Genetic organization and nucleotide sequence of the stability locus
RT   of IncFII plasmid NR1.";
RL   J. Mol. Biol. 202:511-525(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   PLASMID=IncFII R1;
RX   MEDLINE=87060986; PubMed=3023637;
RA   Gerdes K., Molin S.;
RT   "Partitioning of plasmid R1. Structural and functional analysis of
RT   the parA locus.";
RL   J. Mol. Biol. 190:269-279(1986).
RN   [3]
RP   REVISIONS TO 39 AND 60.
```

```
RC   PLASMID=IncFII R1;
RA   Gerdes K.;
RL   Submitted (FEB-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS THOUGHT TO BE INVOLVED IN THE
CC       CONTROL OF PLASMID PARTITION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12777; CAA31264.1; -.
DR   EMBL; X04268; CAA27818.1; -.
DR   PIR; S01774; S01774.
DR   PIR; A24920; A24920.
KW   Plasmid; Plasmid partition.
SQ   SEQUENCE   320 AA;  35765 MW;  52B3239C0A567312 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     386 FFKTNNS 392
           |||||||
Db     301 FFKTNNS 307


RESULT 23
YA0D_SCHPO
ID   YA0D_SCHPO     STANDARD;      PRT;   346 AA.
AC   Q09681;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative glycosyl transferase C5H10.13c in chromosome I (EC 2.-.-.-).
GN   SPAC5H10.13C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 34.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49811; CAA89963.1; -.
KW   Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane;
KW   Signal-anchor; Golgi stack; Glycoprotein.
FT   DOMAIN        1     11       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     12     32       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       33    346       LUMENAL (POTENTIAL).
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    142    142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    224    224       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   346 AA;  39107 MW;  1AD26CCDFABC0D3E CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     556 TTFHEAS 562
           |||||||
Db      81 TTFHEAS 87


RESULT 24
AROG_ECOLI
ID   AROG_ECOLI     STANDARD;      PRT;   350 AA.
AC   P00886;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive
DE   (EC 4.1.2.15) (Phospho-2-keto-3-deoxyheptonate aldolase) (DAHP
DE   synthetase) (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase).
GN   AROG OR B0754 OR Z0924 OR ECS0782.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=82274236; PubMed=6125934;
RA   Davies W.D., Davidson B.E.;
RT   "The nucleotide sequence of aroG, the gene for 3-deoxy-D-
RT   arabinoheptulosonate-7-phosphate synthetase (phe) in Escherichia coli
RT   K12.";
RL   Nucleic Acids Res. 10:4045-4058(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
```

```
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   SEQUENCE OF 1-12.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE=99354419; PubMed=10425687;
RA   Shumilin I.A., Kretsinger R.H., Bauerle R.H.;
RT   "Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-
RT   heptulosonate-7-phosphate synthase from Escherichia coli.";
RL   Structure 7:865-875(1999).
CC   -!- FUNCTION: STEREOSPECIFIC CONDENSATION OF PHOSPHOENOLPYRUVATE (PEP)
CC       AND D-ERYTHROSE-4-PHOSPHATE (E4P) GIVING RISE TO 3-DEOXY-D-
CC       ARABINO-HEPTULOSONATE-7-PHOSPHATE (DAHP).
CC   -!- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-arabino-heptonate 7-
CC       phosphate + phosphate = phosphoenolpyruvate + D-erythrose 4-
CC       phosphate + H(2)O.
CC   -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC       first step.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- MISCELLANEOUS: THERE ARE 3 DAHP SYNTHASES, AROG IS
CC       FEEDBACK-INHIBITED BY PHE. THE OTHER 2 DAHP SYNTHASES ARE TYR- AND
CC       TRP-SENSITIVE, RESPECTIVELY.
CC   -!- SIMILARITY: BELONGS TO CLASS-I DAHP SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J01591; AAA23492.1; -.
DR   EMBL; AE000178; AAC73841.1; -.
DR   EMBL; D90714; BAA35416.1; -.
DR   EMBL; AE005253; AAG55083.1; -.
DR   EMBL; AP002553; BAB34205.1; -.
DR   PIR; A01106; ADECHF.
DR   PDB; 1QR7; 31-AUG-99.
DR   SWISS-2DPAGE; P00886; COLI.
DR   EcoGene; EG10079; aroG.
DR   InterPro; IPR001785; DAHP_synth1.
DR   Pfam; PF00793; DAHP_synth_1; 1.
DR   ProDom; PD005060; DAHP_synth1; 1.
DR   TIGRFAMs; TIGR00034; aroFGH; 1.
KW   Aromatic amino acid biosynthesis; Lyase; 3D-structure;
KW   Complete proteome.
SQ   SEQUENCE   350 AA;  38009 MW;  7477D361962E8710 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 350;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      97 AGEFLDM 103
           |||||||
Db     141 AGEFLDM 147


RESULT 25
AROG_HAEIN
ID   AROG_HAEIN     STANDARD;      PRT;   362 AA.
AC   P44303;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 4.1.2.15) (Phospho-
DE   2-keto-3-deoxyheptonate aldolase) (DAHP synthetase) (3-deoxy-
DE   D-arabino-heptulosonate 7-phosphate synthase).
GN   AROG OR HI1547.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: STEREOSPECIFIC CONDENSATION OF PHOSPHOENOLPYRUVATE (PEP)
CC       AND D-ERYTHROSE-4-PHOSPHATE (E4P) GIVING RISE TO 3-DEOXY-D-
CC       ARABINO-HEPTULOSONATE-7-PHOSPHATE (DAHP).
CC   -!- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-arabino-heptonate 7-
CC       phosphate + phosphate = phosphoenolpyruvate + D-erythrose 4-
CC       phosphate + H(2)O.
CC   -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC       first step.
CC   -!- SIMILARITY: BELONGS TO CLASS-I DAHP SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32830; AAC23197.1; -.
DR   HSSP; P00886; 1QR7.
DR   TIGR; HI1547; -.
```

```
DR   InterPro; IPR001785; DAHP_synth1.
DR   Pfam; PF00793; DAHP_synth_1; 1.
DR   ProDom; PD005060; DAHP_synth1; 1.
DR   TIGRFAMs; TIGR00034; aroFGH; 1.
KW   Aromatic amino acid biosynthesis; Lyase; Complete proteome.
SQ   SEQUENCE   362 AA;  39936 MW;  533B76F7B379CD2D CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 362;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      97 AGEFLDM 103
           |||||||
Db     148 AGEFLDM 154


RESULT 26
AKR_ARATH
ID   AKR_ARATH      STANDARD;      PRT;   439 AA.
AC   Q05753;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ankyrin repeat protein (AKRP).
GN   AKR.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. C24; TISSUE=Leaf;
RX   MEDLINE=93104681; PubMed=1281700;
RA   Zhang H., Scheirer D.C., Fowle W.H., Goodman H.M.;
RT   "Expression of antisense or sense RNA of an ankyrin repeat-containing
RT   gene blocks chloroplast differentiation in Arabidopsis.";
RL   Plant Cell 4:1575-1588(1992).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE TEMPORAL AND SPATIAL REGULATION
CC       OF CHLOROPLAST DEVELOPMENT FROM PROPLASTID.
CC   -!- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION OCCURS IN TWO-WEEK-OLD
CC       PLANTS AND DECLINES AS PLANTS DEVELOP FURTHER.
CC   -!- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M82883; AAA32812.1; -.
DR   PIR; JQ1729; JQ1729.
DR   HSSP; P80144; 2MYO.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; ank; 4.
DR   SMART; SM00248; ANK; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   Cytoskeleton; Repeat; ANK repeat; Multigene family.
FT   REPEAT      288    317       ANK 1.
FT   REPEAT      321    350       ANK 2.
FT   REPEAT      354    383       ANK 3.
FT   REPEAT      387    416       ANK 4.
SQ   SEQUENCE   439 AA;  49150 MW;  C371A90028B25BF3 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 439;
  Best Local Similarity   100.0%;  Pred. No. 55;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     829 SLLLPGT 835
           |||||||
Db      40 SLLLPGT 46


RESULT 27
CBPA_SYNP7
ID   CBPA_SYNP7     STANDARD;      PRT;   450 AA.
AC   P39660;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carotenoid-binding protein A precursor.
GN   CBPA.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89255122; PubMed=2498292;
RA   Reddy K.J., Masamoto K., Sherman D.M., Sherman L.A.;
RT   "DNA sequence and regulation of the gene (cbpA) encoding the 42-
RT   kilodalton cytoplasmic membrane carotenoprotein of the cyanobacterium
RT   Synechococcus sp. strain PCC 7942.";
RL   J. Bacteriol. 171:3486-3493(1989).
CC   -!- FUNCTION: CAROTENOPROTEIN.
CC   -!- SUBCELLULAR LOCATION: Inner membrane-associated.
CC   -!- INDUCTION: BY CARBON DIOXIDE-LIMITED CONDITIONS.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: HIGH, TO NITRATE TRANSPORT PROTEIN NRTA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27055; AAA27303.1; -.
DR   PIR; A44751; A44751.
KW   Transport; Inner membrane; Signal.
FT   SIGNAL        1     49       POTENTIAL.
FT   CHAIN        50    450       CAROTENOID-BINDING PROTEIN A.
SQ   SEQUENCE   450 AA;  49090 MW;  BAE81B5CB7596D40 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     460 TAAGASD 466
           |||||||
Db      47 TAAGASD 53


RESULT 28
TBA2_NEUCR
ID   TBA2_NEUCR     STANDARD;      PRT;   451 AA.
AC   P38669;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tubulin alpha-B chain.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ST. LAWRENCE 74 / SL 74 / ORS 6A;
RX   MEDLINE=97306636; PubMed=9163903;
RA   Monnat J., Ortega Perez R., Turian G.;
RT   "Molecular cloning and expression studies of two divergent
RT   alpha-tubulin genes in Neurospora crassa.";
RL   FEMS Microbiol. Lett. 150:33-41(1997).
```

```
CC   -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC       BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC       CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC   -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79404; CAA55941.1; -.
DR   PIR; S45051; S45051.
DR   InterPro; IPR000217; Tubulin.
DR   InterPro; IPR003008; Tubulin_FtsZ.
DR   Pfam; PF00091; tubulin; 1.
DR   PRINTS; PR01161; TUBULIN.
DR   PROSITE; PS00227; TUBULIN; 1.
KW   Microtubules; GTP-binding; Multigene family.
FT   NP_BIND     142    148       GTP (POTENTIAL).
FT   SITE        451    451       INVOLVED IN POLYMERIZATION.
SQ   SEQUENCE   451 AA;  50254 MW;  CC058D548A6946F5 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 451;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     964 MEGEDVE 970
           |||||||
Db     442 MEGEDVE 448


RESULT 29
RADA_LISMO
ID   RADA_LISMO     STANDARD;      PRT;   457 AA.
AC   Q48761; O86063;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA repair protein radA homolog (DNA repair protein sms homolog).
GN   RADA OR SMS OR SARII OR LMO0233.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LO28 / Serovar 1/2c;
RA   Rouquette C.E., Tascon R., Pellegrini E., Bolla J., Berche P.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2175 / Serovar 4b;
RA   Truong T.K., Kathariou S.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-e / Serovar 1/2a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- FUNCTION: MAY PLAY A ROLE IN THE REPAIR OF ENDOGENOUS ALKYLATION
CC       DAMAGE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RECA FAMILY. RADA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U40604; AAC44447.1; -.
DR   EMBL; AF083254; AAC33293.1; -.
DR   EMBL; AL591974; CAD00760.1; -.
DR   ListiList; LMO00233; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR001984; Lon_endopep.
DR   InterPro; IPR004504; Sms.
DR   PRINTS; PR00830; ENDOLAPTASE.
DR   ProDom; PD000229; RecA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR00416; sms; 1.
DR   PROSITE; PS50162; RECA_2; 1.
KW   DNA repair; ATP-binding; Zinc-finger; DNA-binding; Complete proteome.
FT   ZN_FING      12     29       C4-TYPE (POTENTIAL).
FT   NP_BIND      97    104       ATP (POTENTIAL).
FT   CONFLICT     37     37       E -> K (IN REF. 2).
FT   CONFLICT     73     73       E -> Q (IN REF. 2).
FT   CONFLICT     75     75       N -> S (IN REF. 2).
FT   CONFLICT     86     86       G -> S (IN REF. 2).
FT   CONFLICT    132    132       Q -> R (IN REF. 2).
FT   CONFLICT    181    181       S -> P (IN REF. 1 AND 2).
FT   CONFLICT    197    197       T -> S (IN REF. 2).
FT   CONFLICT    226    227       PR -> AP (IN REF. 2).
FT   CONFLICT    244    244       H -> P (IN REF. 2).
FT   CONFLICT    256    256       R -> P (IN REF. 2).
FT   CONFLICT    260    261       TN -> PI (IN REF. 2).
FT   CONFLICT    290    290       G -> S (IN REF. 2).
FT   CONFLICT    293    293       G -> R (IN REF. 2).
FT   CONFLICT    299    317       SMEGTRPVLVEIQALVSPT -> LWKNSPRPCGNTSACFAN
FT                                (IN REF. 1).
FT   CONFLICT    403    403       R -> C (IN REF. 2).
FT   CONFLICT    404    404       R -> S (IN REF. 1).
FT   CONFLICT    424    424       F -> S (IN REF. 1).
FT   CONFLICT    452    455       MISSING (IN REF. 1).
SQ   SEQUENCE   457 AA;  49981 MW;  8F7386FC40131359 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 57;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     172 ECTATLM 178
           |||||||
Db     195 ECTATLM 201


RESULT 30
CAB1_METJA
ID   CAB1_METJA     STANDARD;      PRT;   482 AA.
AC   Q58773;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carbamoyl-phosphate synthase large chain, N-terminal section (EC
DE   6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN   CARB1 OR MJ1378.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC       phosphate + L-glutamate + carbamoyl phosphate.
CC   -!- COFACTOR: Binds three manganese ions (By similarity).
CC   -!- PATHWAY: Arginine biosynthesis.
CC   -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC   -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC       promotes the hydrolysis of glutamine to ammonia, which is used by
CC       the large (or ammonia) chain to synthesize carbamoyl phosphate.
CC   -!- SIMILARITY: BELONGS TO THE CARB FAMILY. N-TERMINAL SECTION.
CC   -!- CAUTION: Sequence of carB is split into two genes in M.jannaschii
CC       (MJ1378 and MJ1381).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67577; AAB99385.1; -.
DR   HSSP; P00968; 1JDB.
DR   TIGR; MJ1378; -.
DR   InterPro; IPR005483; CPase_L.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR005480; CPase_L_D3.
DR   InterPro; IPR005481; CPase_L_N.
DR   Pfam; PF00289; CPSase_L_chain; 1.
DR   Pfam; PF02786; CPSase_L_D2; 1.
DR   Pfam; PF02787; CPSase_L_D3; 1.
DR   PRINTS; PR00098; CPSASE.
DR   PROSITE; PS00866; CPSASE_1; 1.
DR   PROSITE; PS00867; CPSASE_2; 1.
KW   Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; ATP-binding;
KW   Manganese; Complete proteome.
FT   DOMAIN        1    398       CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT   NP_BIND     150    207       ATP (POTENTIAL).
FT   NP_BIND     299    349       ATP (POTENTIAL).
FT   METAL       281    281       MANGANESE 1 (BY SIMILARITY).
FT   METAL       295    295       MANGANESE 1 AND 2 (BY SIMILARITY).
FT   METAL       297    297       MANGANESE 2 (BY SIMILARITY).
SQ   SEQUENCE   482 AA;  53608 MW;  A8ABE2570B2E3D96 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 482;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     689 TSKGLKY 695
           |||||||
Db     189 TSKGLKY 195


RESULT 31
PEN3_ADEG1
ID   PEN3_ADEG1     STANDARD;      PRT;   515 AA.
AC   Q64755;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Penton protein (Virion component III) (Penton base protein).
GN   PIII.
OS   Avian adenovirus gal1 (strain Phelps) (Fowl adenovirus 1) (CELO).
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX   NCBI_TaxID=10553;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96186720; PubMed=8627769;
RA   Chiocca S., Kurzbauer R., Schaffner G., Baker A., Mautner V.,
RA   Cotten M.;
RT   "The complete DNA sequence and genomic organization of the avian
RT   adenovirus CELO.";
RL   J. Virol. 70:2939-2949(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U46933; AAC54908.1; -.
DR   InterPro; IPR002605; Adeno_Penton_B.
DR   Pfam; PF01686; Adeno_Penton_B; 1.
KW   Late protein.
SQ   SEQUENCE   515 AA;  56722 MW;  90C389ACD686C6AC CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 515;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     720 DLRIPEG 726
           |||||||
Db     181 DLRIPEG 187


RESULT 32
RO60_HUMAN
ID   RO60_HUMAN     STANDARD;      PRT;   538 AA.
AC   P10155; Q92787;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60-kDa SS-A/Ro ribonucleoprotein (60 kDa Ro protein) (60 kDa
DE   ribonucleoprotein Ro) (RoRNP) (Ro 60 kDa autoantigen) (Sjogren
DE   syndrome type A antigen) (SS-A) (Sjogren syndrome antigen A2).
GN   SSA2 OR RO60.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89071722; PubMed=3200833;
RA   Deutscher S.L., Harley J.B., Keene J.D.;
RT   "Molecular analysis of the 60-kDa human Ro ribonucleoprotein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:9479-9483(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89198084; PubMed=2649513;
RA   Ben-Chetrit E., Gandy B.J., Tan E.M., Sulivan K.F.;
RT   "Isolation and characterization of a cDNA clone encoding the 60-kD
RT   component of the human SS-A/Ro ribonucleoprotein autoantigen.";
RL   J. Clin. Invest. 83:1284-1292(1989).
RN   [3]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RA   Buyon J.P., DiDonato F., Tseng C.E., Rashbaum W., Morris A.,
RA   Hamel J.C., Chan E.K.L.;
RT   "Identification and characterization of an alternative mRNA transcript
RT   of the 60-kD SS-A/Ro ribonucleoprotein encoding the N-terminal RNA
RT   binding domain alone.";
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: RNA-BINDING PROTEIN THAT BINDS TO SEVERAL SMALL
CC       CYTOPLASMIC RNA MOLECULES KNOWN AS Y RNAS. MAY STABILIZE THESE
CC       RNAS FROM DEGRADATION.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM/60E2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DISEASE: SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS
CC       OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR
CC       RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
CC   -!- SIMILARITY: BELONGS TO THE RO 60 KDA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04137; AAA35493.1; -.
DR   EMBL; M25077; AAA35532.1; -.
DR   EMBL; U44388; AAB81552.1; ALT_TERM.
DR   EMBL; U44388; AAB81553.1; -.
DR   PIR; A31760; A31760.
DR   PIR; A30596; A30596.
DR   Genew; HGNC:11313; SSA2.
DR   MIM; 600063; -.
DR   MIM; 234700; -.
DR   InterPro; IPR000504; RNA_rec_mot.
KW   Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen;
KW   Alternative splicing.
FT   VARSPLIC    195    205       LAIVTKYITKG -> KHKIFIGKKGG (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC    206    538       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    239    239       K -> R (IN REF. 2).
FT   CONFLICT    515    538       GMLDMCGFDTGALDVIRNFTLDMI -> ALQNTLLNKSF
FT                                (IN REF. 2).
SQ   SEQUENCE   538 AA;  60642 MW;  B9C58AF3248DC48C CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 538;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     769 VTTDMTL 775
           |||||||
Db     424 VTTDMTL 430


RESULT 33
RO60_MOUSE
ID   RO60_MOUSE     STANDARD;      PRT;   538 AA.
AC   O08848; Q9QYD8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60-kDa SS-A/Ro ribonucleoprotein (60 kDa Ro protein) (60 kDa
DE   ribonucleoprotein Ro) (RoRNP).
GN   SSA2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=97266462; PubMed=9112230;
RA   Wang D., Buyon J.P., Chan E.K.L.;
RT   "Cloning and expression of mouse 60 kDa ribonucleoprotein SS-A/Ro.";
RL   Mol. Biol. Rep. 23:205-210(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Kaufman K.M., Farris A.D., Gross J.K., Harley J.B.;
RT   "Characterization of the murine 60 kD Ro gene: genomic sequence,
RT   organization, and chromosomal localization.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 82-538 FROM N.A.
RA   Scofield R.H., Kurien B.T., Kaufman K.M., James J.A., Baber U.,
RA   Harley J.B.;
RT   "Immunization of mice with human 60 kd Ro peptides results in epitope
RT   spreading if the peptides are highly homologous between man and
RT   mouse.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RNA-BINDING PROTEIN THAT BINDS TO SEVERAL SMALL
CC       CYTOPLASMIC RNA MOLECULES KNOWN AS Y RNAS. MAY STABILIZE THESE
CC       RNAS FROM DEGRADATION.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: HIGHEST IN BRAIN, FOLLOWED BY LUNG, MUSCLE,
CC       KIDNEY AND HEART. LOWER LEVELS ARE FOUND IN TESTIS, LIVER AND
CC       SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE RO 60 KDA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66843; AAC53142.1; -.
DR   EMBL; AF065398; AAF19049.1; -.
DR   EMBL; AF042139; AAC15667.1; -.
DR   MGD; MGI:106652; Ssa2.
KW   Ribonucleoprotein; RNA-binding.
FT   CONFLICT     32     33       RL -> V (IN REF. 2).
FT   CONFLICT    458    458       D -> G (IN REF. 2).
FT   CONFLICT    465    465       V -> I (IN REF. 2).
SQ   SEQUENCE   538 AA;  60123 MW;  7850DA35D1726BDA CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 538;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     769 VTTDMTL 775
           |||||||
Db     424 VTTDMTL 430


RESULT 34
GV7_XENLA
ID   GV7_XENLA      STANDARD;      PRT;   591 AA.
AC   P20398;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-NOV-1991 (Rel. 20, Last annotation update)
DE   Developmental protein xLGV7.
GN   GV7.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89252831; PubMed=2721962;
RA   Miller M., Kloc M., Reddy B., Eastman E., Dreyer C., Etkin L.;
RT   "xlgv7: a maternal gene product localized in nuclei of the central
RT   nervous system in Xenopus laevis.";
RL   Genes Dev. 3:572-583(1989).
CC   -!- FUNCTION: MATERNAL PROTEIN THAT MAY HAVE MULTIPLE FUNCTIONS
CC       DURING DEVELOPMENT, ONE OF WHICH MAY BE ASSOCIATED WITH THE
CC       DEVELOPMENT AND MAINTENANCE OF THE CENTRAL NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (FOLLOWING OOCYTE MATURATION),
CC       THEN NUCLEAR (BLASTULA/GASTRULA STAGE).
```

```
CC   -!- TISSUE SPECIFICITY: IN XENOPUS OOCYTE, IN THE CENTRAL NERVOUS
CC       SYSTEM CELLS OF TADPOLES AND ADULT FROGS, AND TRANSIENTLY IN
CC       EPITHELIAL CELLS OF STOMACH AND GUT OF TADPOLES.
CC   -!- DEVELOPMENTAL STAGE: NEURULA STAGE AND IN ADULT BRAIN.
CC   -!- PTM: MAY BE PHOSPHORYLATED DURING OOCYTE MATURATION.
CC   -!- SIMILARITY: CLOSELY RELATED OR IDENTICAL TO NUCLEAR PROTEIN 7.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15627; CAA33637.1; -.
DR   PIR; A30098; A30098.
DR   InterPro; IPR001230; Prenyl_site.
KW   Nuclear protein; Developmental protein; DNA-binding; Phosphorylation;
KW   ATP-binding; Brain; Differentiation; Repeat.
FT   NP_BIND     120    129       ATP (POTENTIAL).
FT   DOMAIN      579    583       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      101    437       HIS-RICH AND REPEAT-RICH DOMAIN, POSSIBLE
FT                                DNA-BINDING REGION.
FT   REPEAT      171    174
FT   REPEAT      183    186
FT   REPEAT      224    227
FT   REPEAT      234    237
SQ   SEQUENCE   591 AA;  66174 MW;  9F3364CE52B3B540 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 591;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     739 CQAVIIP 745
           |||||||
Db     509 CQAVIIP 515


RESULT 35
MTHR_SCHPO
ID   MTHR_SCHPO     STANDARD;      PRT;   603 AA.
AC   Q10258;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable methylenetetrahydrofolate reductase 1 (EC 1.5.1.20).
GN   SPAC56F8.10.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- CATALYTIC ACTIVITY: 5-methyltetrahydrofolate + NADP(+) = 5,10-
CC       methylenetetrahydrofolate + NADPH.
CC   -!- COFACTOR: FAD (BY SIMILARITY).
CC   -!- PATHWAY: Folate metabolism.
CC   -!- SIMILARITY: BELONGS TO THE METHYLENETETRAHYDROFOLATE REDUCTASE
CC       (EC 1.5.1.20/EC 1.7.99.5) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69728; CAA93581.1; -.
DR   HSSP; P00394; 1B5T.
DR   InterPro; IPR004621; Fadh2_euk.
DR   InterPro; IPR003171; Mehydrof_redctse.
DR   Pfam; PF02219; MTHFR; 1.
DR   TIGRFAMs; TIGR00677; fadh2_euk; 1.
KW   Oxidoreductase; Flavoprotein; FAD; NADP.
SQ   SEQUENCE   603 AA;  69012 MW;  38519FEE783D331F CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 603;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     127 WDELPHG 133
           |||||||
Db     342 WDELPHG 348


RESULT 36
CAB2_METJA
ID   CAB2_METJA     STANDARD;      PRT;   618 AA.
AC   Q58776;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carbamoyl-phosphate synthase large chain, C-terminal section (EC
DE   6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
GN   CARB2 OR MJ1381.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
```

```
CC       phosphate + L-glutamate + carbamoyl phosphate.
CC   -!- COFACTOR: Binds three manganese ions (By similarity).
CC   -!- PATHWAY: Arginine biosynthesis.
CC   -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC   -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC       promotes the hydrolysis of glutamine to ammonia, which is used by
CC       the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE CARB FAMILY. C-TERMINAL SECTION.
CC   -!- CAUTION: Sequence of carB is split into two genes in M.jannaschii
CC       (MJ1378 and MJ1381).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67578; AAB99391.1; -.
DR   HSSP; P00968; 1BXR.
DR   TIGR; MJ1381; -.
DR   InterPro; IPR005483; CPase_L.
DR   InterPro; IPR005479; CPase_L_D2.
DR   InterPro; IPR005480; CPase_L_D3.
DR   InterPro; IPR005481; CPase_L_N.
DR   InterPro; IPR004362; MGS_like.
DR   Pfam; PF00289; CPSase_L_chain; 1.
DR   Pfam; PF02786; CPSase_L_D2; 2.
DR   Pfam; PF02787; CPSase_L_D3; 1.
DR   Pfam; PF02142; MGS; 1.
DR   PRINTS; PR00098; CPSASE.
DR   PROSITE; PS00866; CPSASE_1; 1.
DR   PROSITE; PS00867; CPSASE_2; 1.
KW   Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; ATP-binding;
KW   Manganese; Complete proteome.
FT   DOMAIN       81    477       CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT   DOMAIN      478    618       ALLOSTERIC DOMAIN.
FT   METAL       358    358       MANGANESE 3 (BY SIMILARITY).
FT   METAL       370    370       MANGANESE 3 (BY SIMILARITY).
SQ   SEQUENCE   618 AA;  68453 MW;  8C3D4D9C192A954E CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 618;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     137 LSANMEL 143
           |||||||
Db     472 LSANMEL 478


RESULT 37
NKX3_RAT
ID   NKX3_RAT       STANDARD;      PRT;   624 AA.
AC   Q9EPQ0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 3) (Fragment).
GN   SLC24A3 OR NCKX3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=21303617; PubMed=11294880;
RA   Kraev A., Quednau B.D., Leach S., Li X.-F., Dong H., Winkfein R.,
RA   Perizzolo M., Cai X., Yang R., Philipson K.D., Lytton J.;
RT   "Molecular cloning of a third member of the potassium-dependent
RT   sodium-calcium exchanger gene family, NCKX3.";
RL   J. Biol. Chem. 276:23161-23172(2001).
CC   -!- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four
CC       Na(+) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Abundant in the brain. Expressed at low levels
CC       in the aorta, uterus and intestine.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY009158; AAG32680.1; -.
DR   InterPro; IPR004481; K_NaCaexchng.
DR   InterPro; IPR004837; NaCa_Exmemb.
DR   Pfam; PF01699; Na_Ca_Ex; 2.
DR   TIGRFAMs; TIGR00367; K_NaCaexchang-rel; 1.
KW   Transport; Antiport; Symport; Calcium transport; Potassium transport;
KW   Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL       <1     25       POTENTIAL.
FT   CHAIN        26    624       SODIUM/POTASSIUM/CALCIUM EXCHANGER 3.
FT   DOMAIN       26     88       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     89    109       POTENTIAL.
FT   DOMAIN      110    133       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    134    154       POTENTIAL.
FT   DOMAIN      155    163       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    164    184       POTENTIAL.
FT   DOMAIN      185    191       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    192    212       POTENTIAL.
FT   DOMAIN      213    215       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    216    236       POTENTIAL.
FT   DOMAIN      237    465       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    466    486       POTENTIAL.
FT   DOMAIN      487    491       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    492    512       POTENTIAL.
FT   DOMAIN      513    530       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    531    551       POTENTIAL.
FT   DOMAIN      552    561       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    562    582       POTENTIAL.
FT   DOMAIN      583    596       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    597    617       POTENTIAL.
FT   DOMAIN      618    624       EXTRACELLULAR (POTENTIAL).
FT   REPEAT      130    170       ALPHA-1.
FT   REPEAT      499    530       ALPHA-2.
FT   CARBOHYD     52     52       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   624 AA;  69388 MW;  ADCA175AF1BF1141 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 624;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     790 LGIPDVI 796
           |||||||
Db     486 LGIPDVI 492


RESULT 38
NKX3_HUMAN
ID   NKX3_HUMAN     STANDARD;      PRT;   644 AA.
AC   Q9HC58; Q9BQJ9; Q9BQL7; Q9BQY3; Q9H519;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 3).
```

```
GN   SLC24A3 OR NCKX3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain, and Skeletal muscle;
RX   MEDLINE=21303617; PubMed=11294880;
RA   Kraev A., Quednau B.D., Leach S., Li X.-F., Dong H., Winkfein R.,
RA   Perizzolo M., Cai X., Yang R., Philipson K.D., Lytton J.;
RT   "Molecular cloning of a third member of the potassium-dependent
RT   sodium-calcium exchanger gene family, NCKX3.";
RL   J. Biol. Chem. 276:23161-23172(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
CC   -!- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four
CC       Na(+) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Abundant in the brain. Expressed at low levels
CC       in the aorta, uterus and intestine.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF288087; -; NOT_ANNOTATED_CDS.
DR   EMBL; AF169257; AAG12988.2; -.
DR   EMBL; AL121830; CAC36052.1; -.
DR   EMBL; AL049647; CAC36050.1; -.
DR   EMBL; AL034425; CAC36051.1; -.
DR   EMBL; AL139429; CAC13155.1; -.
DR   Genew; HGNC:10977; SLC24A3.
DR   InterPro; IPR004481; K_NaCaexchng.
DR   InterPro; IPR004837; NaCa_Exmemb.
DR   Pfam; PF01699; Na_Ca_Ex; 4.
DR   TIGRFAMs; TIGR00367; K_NaCaexchang-rel; 1.
KW   Transport; Antiport; Symport; Calcium transport; Potassium transport;
KW   Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat;
KW   Alternative splicing.
FT   SIGNAL       1    44       POTENTIAL.
FT   CHAIN       45   644       SODIUM/POTASSIUM/CALCIUM EXCHANGER 3.
FT   DOMAIN      45   107       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   108   128       POTENTIAL.
FT   DOMAIN     129   153       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   154   174       POTENTIAL.
FT   DOMAIN     175   182       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   183   203       POTENTIAL.
FT   DOMAIN     204   210       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   211   231       POTENTIAL.
FT   DOMAIN     232   234       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   235   255       POTENTIAL.
FT   DOMAIN     256   484       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   485   505       POTENTIAL.
FT   DOMAIN     506   510       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   511   531       POTENTIAL.
FT   DOMAIN     532   549       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   550   570       POTENTIAL.
FT   DOMAIN     571   580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   581   601       POTENTIAL.
FT   DOMAIN     602   615       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   616   636       POTENTIAL.
FT   DOMAIN     637   644       EXTRACELLULAR (POTENTIAL).
FT   REPEAT     149   189       ALPHA-1.
FT   REPEAT     518   549       ALPHA-2.
FT   DOMAIN     422   428       POLY-GLU.
FT   CARBOHYD    71    71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    86    86       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   401   441       MISSING (IN ISOFORM 2).
FT   VARSPLIC   442   442       S -> A (IN ISOFORM 2).
FT   CONFLICT    55    55       I -> V (IN REF. 1; AAG12988 AND 2).
SQ   SEQUENCE   644 AA;  72006 MW;  E07EEBB34EFAB56D CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 644;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     790 LGIPDVI 796
           |||||||
Db     505 LGIPDVI 511


RESULT 39
NKX3_MOUSE
ID   NKX3_MOUSE     STANDARD;      PRT;   645 AA.
AC   Q99PD7; Q99PD8; Q99JR2;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/potassium/calcium exchanger 3 precursor (Na(+)/K(+)/Ca(2+)-
DE   exchange protein 3).
GN   SLC24A3 OR NCKX3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ, and CD-1; TISSUE=Embryonic stem cells, and Brain;
RX   MEDLINE=21303617; PubMed=11294880;
RA   Kraev A., Quednau B.D., Leach S., Li X.-F., Dong H., Winkfein R.,
RA   Perizzolo M., Cai X., Yang R., Philipson K.D., Lytton J.;
RT   "Molecular cloning of a third member of the potassium-dependent
RT   sodium-calcium exchanger gene family, NCKX3.";
RL   J. Biol. Chem. 276:23161-23172(2001).
RN   [2]
RP   SEQUENCE OF 323-645 FROM N.A.
RC   TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transports one Ca(2+) and one K(+) in exchange for four
```

```
CC       Na(+) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Abundant in the brain. Highest levels found in
CC       selected thalamic nuclei, hippocampal CA1 neurons and in layer IV
CC       of the cerebral cortex.
CC   -!- SIMILARITY: BELONGS TO THE SLC24A FAMILY OF TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF314821; AAG60049.1; -.
DR   EMBL; AF314822; AAG60050.1; -.
DR   EMBL; BC005742; AAH05742.1; -.
DR   MGD; MGI:2137513; Slc24a3.
DR   InterPro; IPR004481; K_NaCaexchng.
DR   InterPro; IPR004837; NaCa_Exmemb.
DR   Pfam; PF01699; Na_Ca_Ex; 3.
DR   TIGRFAMs; TIGR00367; K_NaCaexchang-rel; 1.
KW   Transport; Antiport; Symport; Calcium transport; Potassium transport;
KW   Sodium transport; Transmembrane; Glycoprotein; Signal; Repeat.
FT   SIGNAL       1     43       POTENTIAL.
FT   CHAIN        44    645      SODIUM/POTASSIUM/CALCIUM EXCHANGER 3.
FT   DOMAIN       44    106      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     107   127      POTENTIAL.
FT   DOMAIN       128   151      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     152   172      POTENTIAL.
FT   DOMAIN       173   181      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     182   202      POTENTIAL.
FT   DOMAIN       203   209      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     210   230      POTENTIAL.
FT   DOMAIN       231   234      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     235   255      POTENTIAL.
FT   DOMAIN       256   486      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     487   507      POTENTIAL.
FT   DOMAIN       508   512      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     513   533      POTENTIAL.
FT   DOMAIN       534   551      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     552   572      POTENTIAL.
FT   DOMAIN       573   582      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     583   603      POTENTIAL.
FT   DOMAIN       604   617      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     618   638      POTENTIAL.
FT   DOMAIN       639   645      EXTRACELLULAR (POTENTIAL).
FT   REPEAT       148   188      ALPHA-1.
FT   REPEAT       520   551      ALPHA-2.
FT   DOMAIN       424   430      POLY-GLU.
FT   CARBOHYD     70    70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85    85       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   645 AA;  71914 MW;  C62DEB6CB4A01C96 CRC64;

  Query Match              0.7%;  Score 7;  DB 1;  Length 645;
  Best Local Similarity  100.0%;  Pred. No. 78;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     790 LGIPDVI 796
           |||||||
Db     507 LGIPDVI 513


RESULT 40
CFPA_TREPA
ID   CFPA_TREPA     STANDARD;      PRT;   677 AA.
AC   Q56336;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytoplasmic filament protein A.
GN   CFPA OR TPN83 OR TP0748.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-21 AND 146-161.
RC   STRAIN=Nichols;
RX   MEDLINE=96236033; PubMed=8655496;
RA   You Y., Elmore S., Colton L.L., Mackenzie C., Stoops J.K.,
RA   Weinstock G.M., Norris S.J.;
RT   "Characterization of the cytoplasmic filament protein gene (cfpA) of
RT   Treponema pallidum subsp. pallidum.";
RL   J. Bacteriol. 178:3177-3187(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FILAMENTS THAT RUN THE
CC       LENGTH OF THE ORGANISM JUST UNDERNEATH THE CYTOPLASMIC MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: AN ARRAY OF 4 TO 6 FILAMENTS LIE IN CLOSE
CC       APPOSITION TO THE INNER MEMBRANE AND ARE ALWAYS LOCALIZED DIRECTLY
CC       UNDERNEATH THE CORRESPONDING GROUP OF PERIPLASMIC FLAGELLA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32683; AAB07757.1; -.
DR   EMBL; AE001246; AAC65715.1; -.
DR   TIGR; TP0748; -.
KW   Structural protein; Antigen; Coiled coil; Complete proteome.
FT   INIT_MET     0     0
FT   DOMAIN       23    48       COILED COIL (POTENTIAL).
FT   DOMAIN       289   310      COILED COIL (POTENTIAL).
FT   DOMAIN       396   416      COILED COIL (POTENTIAL).
FT   DOMAIN       521   539      COILED COIL (POTENTIAL).
SQ   SEQUENCE   677 AA;  78408 MW;  7F77A144214B7948 CRC64;

  Query Match              0.7%;  Score 7;  DB 1;  Length 677;
  Best Local Similarity  100.0%;  Pred. No. 81;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     197 DSSIIFE 203
           |||||||
Db     264 DSSIIFE 270


RESULT 41
NAH4_RAT
ID   NAH4_RAT       STANDARD;      PRT;   717 AA.
AC   P26434;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sodium/hydrogen exchanger 4 (Na(+)/H(+) exchanger 4) (NHE-4).
GN   SLC9A4 OR NHE4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Stomach;
RX   MEDLINE=92250539; PubMed=1577762;
RA   Orlowski J., Kandasamy R.A., Shull G.E.;
RT   "Molecular cloning of putative members of the Na/H exchanger gene
RT   family. cDNA cloning, deduced amino acid sequence, and mRNA tissue
RT   expression of the rat Na/H exchanger NHE-1 and two structurally
RT   related proteins.";
RL   J. Biol. Chem. 267:9331-9339(1992).
CC   -!- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED
CC       BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL
CC       CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD
CC       SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL
CC       TRANSDUCTION.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, FOLLOWED BY COLON
CC       AND SMALL INTESTINE. LESSER AMOUNTS WERE FOUND IN KIDNEY, BRAIN,
CC       UTERUS AND SKELETAL MUSCLE.
CC   -!- PTM: PHOSPHORYLATED (POSSIBLE).
CC   -!- SIMILARITY: BELONGS TO THE NA(+)/H(+) EXCHANGER FAMILY.
CC   -!- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINATION OF HYDROPHOBIC
CC       DOMAINS IN THE NA(+)/H(+) EXCHANGERS VARY AMONG AUTHORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M85301; AAA41703.1; -.
DR   PIR; C40204; C40204.
DR   InterPro; IPR000676; NaH_Exchngr.
DR   InterPro; IPR004709; NaH_exchang3.
DR   Pfam; PF00999; Na_H_Exchanger; 1.
DR   PRINTS; PR01084; NAHEXCHNGR.
DR   TIGRFAMs; TIGR00840; b_cpa1; 1.
KW   Transmembrane; Glycoprotein; Transport; Antiport; Sodium transport;
KW   Multigene family; Phosphorylation.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       14     28       A (M1) HYDROPHOBIC.
FT   DOMAIN       29     69       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       70     90       B (M2) HYDROPHOBIC.
FT   DOMAIN       91     94       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     95    114       C (M3) (POTENTIAL).
FT   DOMAIN      115    127       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    128    148       D (M4) (POTENTIAL).
FT   DOMAIN      149    154       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    155    175       E (M5) (POTENTIAL).
FT   DOMAIN      176    191       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    192    216       F (M5A) (POTENTIAL).
FT   DOMAIN      217    225       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    226    247       G (M5B) (POTENTIAL).
FT   DOMAIN      248    269       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    270    290       H (M6) (POTENTIAL).
FT   DOMAIN      291    304       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    305    325       I (M7) (POTENTIAL).
FT   DOMAIN      326    356       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    357    377       J (M8) (POTENTIAL).
FT   DOMAIN      378    384       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    385    405       K (M9) (POTENTIAL).
FT   DOMAIN      406    420       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      421    441       L, HYDROPHOBIC.
FT   DOMAIN      442    450       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    451    471       M (M10) (POTENTIAL).
FT   DOMAIN      472    717       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    342    342       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   717 AA;  81522 MW;  4EFBBEC7D7782753 CRC64;

  Query Match          0.7%;  Score 7;  DB 1;  Length 717;
  Best Local Similarity  100.0%;  Pred. No. 85;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     147 AAESTGN 153
           |||||||
Db     675 AAESTGN 681


RESULT 42
PRIA_ECOLI
ID   PRIA_ECOLI     STANDARD;      PRT;   732 AA.
AC   P17888;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Primosomal protein N' (Replication factor Y).
GN   PRIA OR B3935.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.
RX   MEDLINE=90280427; PubMed=2162050;
RA   Lee E.H., Masai H., Allen G.C. Jr., Kornberg A.;
RT   "The priA gene encoding the primosomal replicative n' protein of
RT   Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4620-4624(1990).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 3-16.
RX   MEDLINE=90280426; PubMed=2162049;
RA   Nurse P., Digate R., Zavitz K., Marians K.;
RT   "Molecular cloning and DNA sequence analysis of Escherichia coli
RT   priA, the gene encoding the primosomal protein replication factor
RT   Y.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:4615-4619(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=93347969; PubMed=8346018;
RA   Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT   region from 87.2 to 89.2 minutes.";
RL   Nucleic Acids Res. 21:3391-3398(1993).
RN   [4]
RP   SIMILARITY TO DEAD-BOX HELICASES.
RX   MEDLINE=92107690; PubMed=1662369;
RA   Ouzounis C.A., Blencowe B.J.;
RT   "Bacterial DNA replication initiation factor priA is related to
RT   proteins belonging to the 'DEAD-box' family.";
RL   Nucleic Acids Res. 19:6953-6953(1991).
CC   -!- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA;
CC       THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND
CC       PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT
CC       ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A
CC       HELICASE WITHIN THE PRIMOSOME.
CC   -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. PRIA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M33293; AAA24416.1; -.
DR   EMBL; D00616; BAA00491.1; -.
DR   EMBL; L19201; AAB03067.1; -.
DR   EMBL; AE000467; AAC76917.1; -.
DR   PIR; A35505; A35505.
DR   PIR; A35506; A35506.
DR   ECO2DBASE; H096.8; 6TH EDITION.
DR   EcoGene; EG10763; priA.
```

```
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR005259; PriA.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   TIGRFAMs; TIGR00595; priA; 1.
KW   DNA replication; DNA-binding; ATP-binding; Helicase; Primosome;
KW   Zinc-finger; Complete proteome.
FT   NP_BIND     224    231       ATP (BY SIMILARITY).
FT   SITE        320    323       DEEH BOX.
FT   ZN_FING     436    448       C4-TYPE (POTENTIAL).
FT   ZN_FING     463    479       C4-TYPE (POTENTIAL).
FT   CONFLICT    156    156       R -> A (IN REF. 3).
FT   CONFLICT    621    621       R -> A (IN REF. 3).
FT   CONFLICT    649    649       A -> V (IN REF. 2).
SQ   SEQUENCE   732 AA;  81825 MW;  37A7B35CBB7F3735 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 732;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      26 WRLLLWA 32
           |||||||
Db      78 WRLLLWA 84


RESULT 43
EF2_PYRAE
ID   EF2_PYRAE      STANDARD;      PRT;   740 AA.
AC   Q8ZZC1;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Elongation factor 2 (EF-2).
GN   FUSA OR PAE0332.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC   -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC       the nascent protein chain from the A-site to the P-site of the
CC       ribosome.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EF-G/EF-2 SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE009762; AAL62720.1; -.
DR   InterPro; IPR004543; aEF-2.
DR   InterPro; IPR000640; EFG_C.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00679; EFG_C; 1.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 1.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   TIGRFAMs; TIGR00490; aEF-2; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00301; EFACTOR_GTP; FALSE_NEG.
KW   Elongation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   NP_BIND      32     39       GTP (BY SIMILARITY).
FT   NP_BIND      98    102       GTP (BY SIMILARITY).
FT   NP_BIND     152    155       GTP (BY SIMILARITY).
FT   MOD_RES     605    605       DIPHTHAMIDE (BY SIMILARITY).
SQ   SEQUENCE   740 AA;  83149 MW;  262E97F71A39CEFF CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 740;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 RIPRLWR 27
           |||||||
Db     270 RIPRLWR 276


RESULT 44
UNC8_CAEEL
ID   UNC8_CAEEL     STANDARD;      PRT;   777 AA.
AC   Q21974;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Degenerin unc-8 (Uncoordinated protein 8).
GN   UNC-8 OR R13A1.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Tavernarakis N., Shreffler W., Wang S.L., Driscoll M.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R., Du Z.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNEL
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U76402; AAB39734.1; -.
DR   EMBL; U40798; AAA81473.2; -.
DR   WormPep; R13A1.4; CE26381.
DR   InterPro; IPR004726; Deg-1.
DR   InterPro; IPR001873; Na+channel_ASC.
DR   Pfam; PF00858; ASC; 1.
DR   PRINTS; PR01078; AMINACHANNEL.
DR   TIGRFAMs; TIGR00867; deg-1; 1.
DR   PROSITE; PS01206; ASC; 1.
KW   Ionic channel; Transmembrane; Ion transport; Glycoprotein.
FT   DOMAIN        1    128       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    129    149       POTENTIAL.
FT   DOMAIN      150    689       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    690    710       POTENTIAL.
FT   DOMAIN      711    777       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    274    274       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    319    319       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    453    453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    533    533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    597    597       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   777 AA;  89127 MW;  9B328C3A24A1E7E6 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 777;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     362 EDLEGAV 368
           |||||||
Db     242 EDLEGAV 248


RESULT 45
RA16_SCHPO
ID   RA16_SCHPO     STANDARD;      PRT;   892 AA.
AC   P36617;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA repair protein rad16.
GN   RAD16 OR RAD10 OR RAD20 OR SWI9 OR SPCC970.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94158877; PubMed=8114734;
RA   Carr A.M., Schmidt H., Kirchoff S., Muriel W.J., Sheldrick K.S.,
RA   Griffiths D.J., Basmacioglu C.N., Subramani S., Clegg M., Nasim A.,
RA   Lehmann A.R.;
RT   "The rad16 gene of Schizosaccharomyces pombe: a homolog of the RAD1
RT   gene of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 14:2029-2040(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: INVOLVED IN NUCLEOTIDE EXCISION REPAIR OF DNA DAMAGED
CC       WITH UV LIGHT, BULKY ADDUCTS, OR CROSS-LINKING AGENTS.
CC       ENDONUCLEASE THAT SPECIFICALLY DEGRADES SINGLE-STRANDED DNA.
CC   -!- SUBUNIT: HETERODIMER COMPOSED OF RAD16 AND SWI10.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE XPF/ERCC4/RAD1/RAD16 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X71595; CAA50599.1; -.
DR   EMBL; AL031530; CAA20694.1; -.
DR   InterPro; IPR004016; ERCC4.
DR   Pfam; PF02732; ERCC4; 1.
DR   TIGRFAMs; TIGR00596; rad1; 1.
KW   DNA repair; DNA-binding; Nuclear protein; Hydrolase; Nuclease;
KW   Endonuclease; Phosphorylation.
FT   MOD_RES      71     71       PHOSPHORYLATION (BY CK2) (POTENTIAL).
FT   VARIANT     674    674       E -> K (IN SWI9 ALLELE).
SQ   SEQUENCE   892 AA;  102073 MW;  6C0F4B51475333AC CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 892;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     826 VPGSLLL 832
           |||||||
Db      47 VPGSLLL 53


RESULT 46
YMH6_YEAST
ID   YMH6_YEAST     STANDARD;      PRT;   944 AA.
AC   Q03631;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative 107.6 kDa transcriptional regulatory protein in CPR3-HMG1
DE   intergenic region.
GN   YML076C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC       CLUSTER DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46373; CAA86502.1; -.
DR   SGD; S0004541; YML076C.
DR   InterPro; IPR001138; Fungi_TrN.
DR   SMART; SM00066; GAL4; 1.
DR   PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR   PROSITE; PS50048; ZN2_CY6_FUNGAL_2; FALSE_NEG.
KW   Hypothetical protein; Transcription regulation; DNA-binding; Zinc;
KW   Nuclear protein; Metal-binding.
FT   DNA_BIND     76    109       ZN(2)-CYS(6), FUNGAL-TYPE.
SQ   SEQUENCE   944 AA;  107560 MW;  020A56745DF52CCC CRC64;
```

```
  Query Match             0.7%;  Score 7;  DB 1;  Length 944;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     264 IAITGVA 270
           |||||||
Db       5 IAITGVA 11


RESULT 47
POP1_HUMAN
ID   POP1_HUMAN     STANDARD;      PRT;  1024 AA.
AC   Q99575; Q15037;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ribonucleases P/MRP protein subunit POP1 (EC 3.1.26.5) (hPOP1).
GN   POP1 OR KIAA0061.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-126 FROM N.A.
RX   MEDLINE=97076154; PubMed=8918471;
RA   Lygerou Z., Pluk H., Van Venrooij W.J., Seraphin B.;
RT   "hPop1: an autoantigenic protein subunit shared by the human RNase P
RT   and RNase MRP ribonucleoproteins.";
RL   EMBO J. 15:5936-5948(1996).
RN   [2]
RP   SEQUENCE OF 122-1024 FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051398; PubMed=7584044;
RA   Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA   Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. II.
RT   The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 1:223-229(1994).
CC   -!- FUNCTION: COMPONENT OF RIBONUCLEASE P, A PROTEIN COMPLEX THAT
CC       GENERATES MATURE TRNA MOLECULES BY CLEAVING THEIR 5' ENDS. ALSO A
CC       COMPONENT OF RNASE MRP.
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC       extra-nucleotide from tRNA precursor.
CC   -!- SUBUNIT: RNASE P CONSISTS OF A RNA MOIETY AND AT LEAST 8 PROTEIN
CC       SUBUNITS; POP1, RPP14, RPP20/POP7, RPP25, RPP29/POP4, RPP30,
CC       RPP38 AND RPP40.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99302; CAA67684.1; -.
DR   EMBL; D31765; BAA06543.1; -.
DR   MIM; 602486; -.
KW   Hydrolase; Nuclear protein; tRNA processing.
SQ   SEQUENCE   1024 AA;  114708 MW;  A1DB872F3B940C02 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1024;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     777 GITSPAE 783
           |||||||
Db     492 GITSPAE 498


RESULT 48
MAN1_RAT
ID   MAN1_RAT       STANDARD;      PRT;  1040 AA.
AC   P21139;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Alpha-mannosidase (EC 3.2.1.24) (Alpha-D-mannoside mannohydrolase)
DE   (AMAN).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=91009139; PubMed=2211613;
RA   Bischoff J., Moremen K., Lodish H.F.;
RT   "Isolation, characterization, and expression of cDNA encoding a rat
RT   liver endoplasmic reticulum alpha-mannosidase.";
RL   J. Biol. Chem. 265:17110-17117(1990).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC       mannose residues in alpha-D-mannosides.
CC   -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC   -!- PTM: A SOLUBLE ISOFORM OF THE RAT LIVER MANNOSIDASE IS SUPPOSED TO
CC       BE DERIVED FROM ENDOPLASMIC RETICULUM MEMBRANE ALPHA-MANNOSIDASE
CC       BY PROTEOLYSIS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57547; AAA41565.1; -.
DR   PIR; A38306; A38306.
DR   InterPro; IPR000602; Glyco_hydro_38.
DR   Pfam; PF01074; Glyco_hydro_38; 1.
KW   Hydrolase; Glycosidase; Endoplasmic reticulum.
SQ   SEQUENCE   1040 AA;  115970 MW;  144F6EB97C8F7EA7 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1040;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 RLWRLLL 30
           |||||||
Db     565 RLWRLLL 571


RESULT 49
POLS_EEVV8
ID   POLS_EEVV8     STANDARD;      PRT;  1254 AA.
AC   P05674;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Structural polyprotein (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE   (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE   6 kDa peptide; Spike glycoprotein E1].
OS   Venezuelan equine encephalitis virus (strain TC-83).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=11037;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86306669; PubMed=3755750;
RA   Johnson B.J.B., Kinney R.M., Kost C.L., Trent D.W.;
RT   "Molecular determinants of alphavirus neurovirulence: nucleotide and
RT   deduced protein sequence changes during attenuation of Venezuelan
RT   equine encephalitis virus.";
```

```
RL   J. Gen. Virol. 67:1951-1960(1986).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04368; CAA27883.1; -.
DR   PIR; A27871; VHWVVE.
DR   HSSP; P03315; 1VCP.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR002548; Alpha_E1_glycop.
DR   InterPro; IPR000936; Alpha_E2_glycop.
DR   InterPro; IPR002533; Alpha_E3_glycop.
DR   InterPro; IPR001836; Alpha_core.
DR   InterPro; IPR000930; Togavirin.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    275       COAT PROTEIN C.
FT   CHAIN       276    334       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       335    757       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       758    812       6 KDA PEPTIDE.
FT   CHAIN       813   1254       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    152    152       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    158    158       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    226    226       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    702    718       POTENTIAL.
FT   TRANSMEM    774    790       POTENTIAL.
FT   TRANSMEM    792    808       POTENTIAL.
FT   TRANSMEM   1231   1247       POTENTIAL.
FT   CARBOHYD    286    286       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    546    546       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    652    652       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    946    946       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1254 AA;  138485 MW;  7615698519A529F6 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    224 TEKGWEF 230
          |||||||
Db    656 TEKGWEF 662


RESULT 50
POLS_EEVVM
ID   POLS_EEVVM     STANDARD;      PRT;  1254 AA.
AC   P36331; Q66587; Q66588; Q66589; Q66590; Q66591;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Structural polyprotein (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE   (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE   6 kDa peptide; Spike glycoprotein E1].
OS   Venezuelan equine encephalitis virus (strain Mena II).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=36384;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93187617; PubMed=8445371;
RA   Sneider J.M., Kinney R.M., Tsuchiya K.R., Trent D.W.;
RT   "Molecular evidence that epizootic Venezuelan equine encephalitis
RT   (VEE) I-AB viruses are not evolutionary derivatives of enzootic VEE
RT   subtype I-E or II viruses.";
RL   J. Gen. Virol. 74:519-523(1993).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04599; AAA42990.1; -.
DR   EMBL; L04599; AAA42989.1; ALT_TERM.
DR   EMBL; L04599; AAA42991.1; ALT_SEQ.
DR   EMBL; L04599; AAA42992.1; ALT_SEQ.
DR   EMBL; L04599; AAA42993.1; ALT_SEQ.
DR   EMBL; L04599; AAA42994.1; ALT_INIT.
DR   PIR; JQ1979; JQ1979.
DR   HSSP; P03315; 1VCP.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR002548; Alpha_E1_glycop.
DR   InterPro; IPR000936; Alpha_E2_glycop.
DR   InterPro; IPR002533; Alpha_E3_glycop.
DR   InterPro; IPR001836; Alpha_core.
DR   InterPro; IPR000930; Togavirin.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    274       COAT PROTEIN C.
FT   CHAIN       275    333       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       334    756       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       757    812       6 KDA PEPTIDE.
FT   CHAIN       813   1254       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    151    151       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    157    157       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    225    225       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    701    721       POTENTIAL.
FT   TRANSMEM    794    813       POTENTIAL.
FT   TRANSMEM   1231   1248       POTENTIAL.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    946    946       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1254 AA;  138343 MW;  FB9DE88265F20211 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    224 TEKGWEF 230
          |||||||
Db    655 TEKGWEF 661


RESULT 51
POLS_EEVVT
```

```
ID   POLS_EEVVT     STANDARD;      PRT;  1254 AA.
AC   P09592; Q88691; Q88692; Q88693; Q88694; Q88695;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Structural polyprotein (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE   (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE   6 kDa peptide; Spike glycoprotein E1].
OS   Venezuelan equine encephalitis virus (strain Trinidad donkey).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=11038;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86263392; PubMed=3088830;
RA   Kinney R.M., Johnson B.J.B., Brown V.L., Trent D.W.;
RT   "Nucleotide sequence of the 26 S mRNA of the virulent Trinidad donkey
RT   strain of Venezuelan equine encephalitis virus and deduced sequence
RT   of the encoded structural proteins.";
RL   Virology 152:400-413(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89243175; PubMed=2524126;
RA   Kinney R.M., Johnson B.J.B., Welch J.B., Tsuchiya K.R., Trent D.W.;
RT   "The full-length nucleotide sequences of the virulent Trinidad donkey
RT   strain of Venezuelan equine encephalitis virus and its attenuated
RT   vaccine derivative, strain TC-83.";
RL   Virology 170:19-30(1989).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14937; AAA42997.1; -.
DR   EMBL; J04332; AAB02519.1; -.
DR   PIR; B31467; VHWVVT.
DR   HSSP; P03315; 1VCP.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR002548; Alpha_E1_glycop.
DR   InterPro; IPR000936; Alpha_E2_glycop.
DR   InterPro; IPR002533; Alpha_E3_glycop.
DR   InterPro; IPR001836; Alpha_core.
DR   InterPro; IPR000930; Togavirin.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    275       COAT PROTEIN C.
FT   CHAIN       276    334       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       335    757       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       758    812       6 KDA PEPTIDE.
FT   CHAIN       813   1254       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    152    152       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    158    158       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    226    226       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    702    718       POTENTIAL.
FT   TRANSMEM    774    790       POTENTIAL.
FT   TRANSMEM    792    808       POTENTIAL.
FT   TRANSMEM   1231   1247       POTENTIAL.
FT   CARBOHYD    286    286       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    546    546       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    652    652       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    946    946       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    811    811       P -> G (IN REF. 2).
SQ   SEQUENCE   1254 AA;  138390 MW;  1C0F62BE1909664B CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     224 TEKGWEF 230
           |||||||
Db     656 TEKGWEF 662


RESULT 52
POLS_EEVV3
ID   POLS_EEVV3     STANDARD;      PRT;  1255 AA.
AC   P36329;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Structural polyprotein (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE   (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE   6 kDa peptide; Spike glycoprotein E1].
OS   Venezuelan equine encephalitis virus (strain 3880).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=36382;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93079859; PubMed=1448915;
RA   Kinney R.M., Tsuchiya K.R., Sneider J.M., Trent D.W.;
RT   "Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)
RT   viruses may have evolved from enzootic VEE subtype I-D virus.";
RL   Virology 191:569-580(1992).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00930; AAC19325.1; -.
DR   PIR; D44213; D44213.
DR   HSSP; P03315; 1VCP.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR002548; Alpha_E1_glycop.
DR   InterPro; IPR000936; Alpha_E2_glycop.
DR   InterPro; IPR002533; Alpha_E3_glycop.
DR   InterPro; IPR001836; Alpha_core.
DR   InterPro; IPR000930; Togavirin.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    275       COAT PROTEIN C.
FT   CHAIN       276    334       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       335    757       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       758    813       6 KDA PEPTIDE.
FT   CHAIN       814   1255       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    152    152       CHARGE RELAY SYSTEM (BY SIMILARITY).
```

```
FT   ACT_SITE    158    158       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    226    226       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    702    722       POTENTIAL.
FT   TRANSMEM    795    814       POTENTIAL.
FT   TRANSMEM   1232   1249       POTENTIAL.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    286    286       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    652    652       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1255 AA;  138297 MW;  7D730E17CAECA310 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     224 TEKGWEF 230
           |||||||
Db     656 TEKGWEF 662


RESULT 53
POLS_EEVVP
ID   POLS_EEVVP     STANDARD;      PRT;  1255 AA.
AC   P36332;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Structural polyprotein (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE   (Capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE   6 kDa peptide; Spike glycoprotein E1].
OS   Venezuelan equine encephalitis virus (strain P676).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=36385;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93079859; PubMed=1448915;
RA   Kinney R.M., Tsuchiya K.R., Sneider J.M., Trent D.W.;
RT   "Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)
RT   viruses may have evolved from enzootic VEE subtype I-D virus.";
RL   Virology 191:569-580(1992).
CC   -!- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC       SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC       HEMAGGLUTININ.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L04653; AAC19319.1; -.
DR   PIR; B44213; B44213.
DR   HSSP; P03315; 1VCP.
DR   MEROPS; S03.001; -.
DR   InterPro; IPR002548; Alpha_E1_glycop.
DR   InterPro; IPR000936; Alpha_E2_glycop.
DR   InterPro; IPR002533; Alpha_E3_glycop.
DR   InterPro; IPR001836; Alpha_core.
DR   InterPro; IPR000930; Togavirin.
DR   Pfam; PF00943; Alpha_E2_glycop; 1.
DR   Pfam; PF00944; Alpha_core; 1.
DR   Pfam; PF01563; Alpha_E3_glycop; 1.
DR   Pfam; PF01589; Alpha_E1_glycop; 1.
DR   PRINTS; PR00798; TOGAVIRIN.
KW   Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
KW   Serine protease.
FT   CHAIN         1    275       COAT PROTEIN C.
FT   CHAIN       276    334       SPIKE GLYCOPROTEIN E3.
FT   CHAIN       335    757       SPIKE GLYCOPROTEIN E2.
FT   CHAIN       758    813       6 KDA PEPTIDE.
FT   CHAIN       814   1255       SPIKE GLYCOPROTEIN E1.
FT   ACT_SITE    152    152       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    158    158       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    226    226       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   TRANSMEM    702    722       POTENTIAL.
FT   TRANSMEM    795    814       POTENTIAL.
FT   TRANSMEM   1232   1249       POTENTIAL.
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    286    286       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    652    652       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    947    947       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1255 AA;  138213 MW;  33CD302F5CAE8646 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     224 TEKGWEF 230
           |||||||
Db     656 TEKGWEF 662


RESULT 54
POLN_SINDO
ID   POLN_SINDO     STANDARD;      PRT;  2514 AA.
AC   P27283;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nonstructural polyprotein (P270) [Contains: Nonstructural protein
DE   NSP1; Nonstructural NSP2; Nonstructural NSP3; Nonstructural NSP4].
OS   Sindbis virus (subtype Ockelbo / strain Edsbyn 82-5).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=31699;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220725; PubMed=1673813;
RA   Shirako Y., Niklasson B., Dalrymple J.M., Strauss E.G., Strauss J.H.;
RT   "Structure of the Ockelbo virus genome and its relationship to other
RT   Sindbis viruses.";
RL   Virology 182:753-764(1991).
CC   -!- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
CC       BETWEEN THE CODONS FOR 1898-TYR AND 1899-LEU.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M69205; AAA96972.1; ALT_SEQ.
DR   PIR; A39991; MNWV82.
DR   MEROPS; C09.001; -.
DR   InterPro; IPR002589; A1pp.
DR   InterPro; IPR002620; Peptidase_C9.
DR   InterPro; IPR001788; RNA_dep_RNApol2.
DR   InterPro; IPR002588; V_methyltransf.
DR   InterPro; IPR000606; Viral_helicase1.
DR   Pfam; PF00978; RNA_dep_RNApol2; 1.
DR   Pfam; PF01443; Viral_helicase1; 1.
DR   Pfam; PF01660; Vmethyltransf; 1.
DR   Pfam; PF01661; A1pp; 1.
DR   Pfam; PF01707; Peptidase_C9; 1.
DR   SMART; SM00506; A1pp; 1.
KW   Polyprotein; Nonstructural protein; RNA-binding; Helicase.
```

```
FT   CHAIN         1    540       NONSTRUCTURAL PROTEIN NSP1.
FT   CHAIN       541   1347       NONSTRUCTURAL PROTEIN NSP2.
FT   CHAIN      1348   1898       NONSTRUCTURAL PROTEIN NSP3.
FT   CHAIN      1899   2514       NONSTRUCTURAL PROTEIN NSP4.
SQ   SEQUENCE   2514 AA;  279642 MW;  2F388CE32ACF5EDD CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 2514;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     532 KGKQSYT 538
           |||||||
Db     268 KGKQSYT 274


RESULT 55
FRAP_HUMAN
ID   FRAP_HUMAN     STANDARD;      PRT;  2549 AA.
AC   P42345; Q9Y4I3;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   FKBP-rapamycin associated protein (FRAP) (Rapamycin target protein).
GN   FRAP1 OR FRAP OR FRAP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=94277209; PubMed=8008069;
RA   Brown E.J., Albers M.W., Shin T.B., Ichikawa K., Keith C.T.,
RA   Lane W.S., Schreiber S.L.;
RT   "A mammalian protein targeted by G1-arresting rapamycin-receptor
RT   complex.";
RL   Nature 369:756-758(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98317532; PubMed=9653645;
RA   Onyango P., Lubyova B., Gardellin P., Kurzbauer R., Weith A.;
RT   "Molecular cloning and expression analysis of five novel genes in
RT   chromosome 1p36.";
RL   Genomics 50:187-198(1998).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 2018-2112.
RX   MEDLINE=96279639; PubMed=8662507;
RA   Choi J., Chen J., Schreiber S.L., Clardy J.;
RT   "Structure of the FKBP12-rapamycin complex interacting with the
RT   binding domain of human FRAP.";
RL   Science 273:239-242(1996).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 2018-2112.
RX   MEDLINE=99190960; PubMed=10089303;
RA   Liang J., Choi J., Clardy J.;
RT   "Refined structure of the FKBP12-rapamycin-FRB ternary complex at 2.2
RT   A resolution.";
RL   Acta Crystallogr. D 55:736-744(1999).
CC   -!- FUNCTION: ACTS AS THE TARGET FOR THE CELL-CYCLE ARREST AND
CC       IMMUNOSUPPRESSIVE EFFECTS OF THE FKBP12-RAPAMYCIN COMPLEX.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   -!- SIMILARITY: CONTAINS 8 HEAT REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34075; AAA58486.1; -.
DR   EMBL; U88966; AAC39933.1; -.
DR   PDB; 1FAP; 23-JUL-97.
DR   PDB; 1NSG; 18-MAR-98.
DR   PDB; 1AUE; 18-NOV-98.
DR   Genew; HGNC:3942; FRAP1.
DR   MIM; 601231; -.
DR   InterPro; IPR003151; FAT.
DR   InterPro; IPR003152; FATC.
DR   InterPro; IPR000357; HEAT_repeat.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   Pfam; PF02259; FAT; 1.
DR   Pfam; PF02260; FATC; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
DR   PROSITE; PS50077; HEAT_REPEAT; FALSE_NEG.
KW   Transferase; Kinase; Repeat; 3D-structure.
FT   REPEAT       16     53       HEAT 1.
FT   REPEAT      650    688       HEAT 2.
FT   REPEAT      859    897       HEAT 3.
FT   REPEAT      988   1025       HEAT 4.
FT   REPEAT     1069   1106       HEAT 5.
FT   REPEAT     1109   1148       HEAT 6.
FT   REPEAT     1150   1186       HEAT 7.
FT   DOMAIN     1382   1982       FAT.
FT   REPEAT     1933   1970       HEAT 8.
FT   DOMAIN     2182   2549       PI3K/PI4K.
FT   CONFLICT    353    353       K -> N (IN REF. 2).
FT   CONFLICT    359    359       S -> N (IN REF. 2).
FT   CONFLICT    364    364       D -> N (IN REF. 2).
FT   CONFLICT    390    390       M -> L (IN REF. 2).
FT   CONFLICT    430    430       R -> L (IN REF. 2).
FT   CONFLICT    455    457       VLD -> GVE (IN REF. 2).
FT   CONFLICT    461    461       A -> G (IN REF. 2).
FT   CONFLICT    482    484       VFT -> FFN (IN REF. 2).
FT   CONFLICT    489    489       L -> V (IN REF. 2).
FT   CONFLICT    513    513       L -> I (IN REF. 2).
FT   CONFLICT    539    539       L -> V (IN REF. 2).
FT   CONFLICT    553    553       R -> C (IN REF. 2).
FT   CONFLICT    956    999       MRIFRDQSLSHHHTMVVQAITFIFKSLGLKCVQFLPQVMPT
FT                                FLN -> ADLPRPVTLSSSHHGCPGHHLHLQVPGTQMCAVP
FT                                APGHAHVPY (IN REF. 2).
FT   CONFLICT   1075   1075       I -> S (IN REF. 2).
SQ   SEQUENCE   2549 AA;  288888 MW;  7D9AD6E784882AB4 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 2549;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     594 EASDVGS 600
           |||||||
Db     575 EASDVGS 581


RESULT 56
7LES_DROME
ID   7LES_DROME     STANDARD;      PRT;  2554 AA.
AC   P13368; Q9U5V7; Q9VZ36; Q9TYI0;
DT   01-JAN-1990 (Rel. 13, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sevenless protein (EC 2.7.1.112).
GN   SEV OR HD-265 OR CG18085.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=88282538; PubMed=2840202;
```

```
RA   Basler K., Hafen E.;
RT   "Control of photoreceptor cell fate by the sevenless protein requires
RT   a functional tyrosine kinase domain.";
RL   Cell 54:299-311(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=88329706; PubMed=3138161;
RA   Bowtell D.L.L., Simon M.A., Rubin G.M.;
RT   "Nucleotide sequence and structure of the sevenless gene of
RT   Drosophila melanogaster.";
RL   Genes Dev. 2:620-634(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE OF 2349-2408 FROM N.A.
RX   MEDLINE=98401146; PubMed=9731193;
RA   Oates A.C., Wollberg P., Achen M.G., Wilks A.F.;
RT   "Sampling the genomic pool of protein tyrosine kinase genes using the
RT   polymerase chain reaction with genomic DNA.";
RL   Biochem. Biophys. Res. Commun. 249:660-667(1998).
RN   [5]
RP   IDENTIFICATION OF FN-III REPEATS.
RX   MEDLINE=90199889; PubMed=2317871;
RA   Norton P.A., Hynes R.O., Ress D.J.G.;
RT   "Sevenless: seven found?";
RL   Cell 61:15-16(1990).
CC   -!- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO
CC       INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE
CC       LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN ON THE
CC       SURFACE OF THE NEIGHBORING R8 CELL.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBUNIT: MAY FORM A COMPLEX WITH DRK AND SOS.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC       RECEPTOR SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION
CC       NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE
CC       NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03158; AAA28882.1; -.
DR   EMBL; X13666; CAA31960.1; ALT_INIT.
DR   EMBL; X13666; CAB55310.1; -.
DR   EMBL; AE003484; AAF47992.1; ALT_INIT.
DR   EMBL; AJ002917; CAA05752.1; -.
DR   PIR; A28912; TVFF7L.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0003366; sev.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000033; LdlI_receptor_rep.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00135; LY; 2.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; Receptor; Vision; Transmembrane;
KW   Glycoprotein; ATP-binding; Phosphorylation; Repeat.
FT   DOMAIN        1   2123       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   2124   2147       POTENTIAL.
FT   DOMAIN     2148   2554       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      311    431       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      436    528       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      822    921       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1298   1392       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1680   1794       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1797   1897       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1898   1988       FIBRONECTIN TYPE-III 7.
FT   DOMAIN     2038   2046       POLY-ARG.
FT   DOMAIN     2209   2485       PROTEIN KINASE.
FT   NP_BIND    2215   2223       ATP (BY SIMILARITY).
FT   BINDING    2242   2242       ATP (BY SIMILARITY).
FT   MOD_RES    2380   2380       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     30     30       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    129    129       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    481    481       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    505    505       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    617    617       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    647    647       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    966    966       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1228   1228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1313   1313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1353   1353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1550   1550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1557   1557       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1639   1639       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1725   1725       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    1756    1756       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1804    1804       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1889    1889       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1947    1947       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2073    2073       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MUTAGEN     2242    2242       K->M: INACTIVATES THE PROTEIN.
FT   CONFLICT     392     392       V -> M (IN REF. 1).
FT   CONFLICT     663     663       A -> T (IN REF. 3).
FT   CONFLICT    1703    1703       N -> H (IN REF. 3).
FT   CONFLICT    1730    1731       RG -> KE (IN REF. 3).
FT   CONFLICT    1741    1741       V -> M (IN REF. 3).
FT   CONFLICT    1823    1823       E -> Q (IN REF. 2).
FT   CONFLICT    2271    2271       C -> R (IN REF. 1).
SQ   SEQUENCE   2554 AA;  287022 MW;  09E238A0F27684F8 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 2554;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     866 AIVSSCV 872
           |||||||
Db    2135 AIVSSCV 2141


RESULT 57
7LES_DROVI
ID   7LES_DROVI     STANDARD;      PRT;  2594 AA.
AC   P20806;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sevenless protein (EC 2.7.1.112).
GN   SEV.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7244;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90319110; PubMed=2115169;
RA   Michael W.M., Bowtell D.D.L., Rubin G.M.;
RT   "Comparison of the sevenless genes of Drosophila virilis and
RT   Drosophila melanogaster.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:5351-5353(1990).
CC   -!- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO
CC       INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE
CC       LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC       RECEPTOR SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION
CC       NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE
CC       NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34545; AAA28883.1; -.
DR   EMBL; M34544; AAA28883.1; JOINED.
DR   EMBL; M34543; AAA28883.1; JOINED.
DR   PIR; A35774; A35774.
DR   HSSP; P11362; 1FGK.
DR   FlyBase; FBgn0013140; Dvir\sev.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR000033; Ldl_receptor_rep.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 5.
DR   SMART; SM00135; LY; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; Receptor; Vision; Transmembrane;
KW   Glycoprotein; ATP-binding; Phosphorylation; Repeat.
FT   DOMAIN         1    2139       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    2140    2163       POTENTIAL.
FT   DOMAIN      2164    2594       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         9      26       POLY-GLN.
FT   DOMAIN       356     459       FIBRONECTIN TYPE-III 1.
FT   DOMAIN       464     555       FIBRONECTIN TYPE-III 2.
FT   DOMAIN       835     935       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      1328    1421       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      1706    1816       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      1817    1916       FIBRONECTIN TYPE-III 6.
FT   DOMAIN      1917    2007       FIBRONECTIN TYPE-III 7.
FT   DOMAIN      2057    2063       POLY-ARG.
FT   DOMAIN      2224    2495       PROTEIN KINASE.
FT   NP_BIND     2230    2238       ATP (BY SIMILARITY).
FT   BINDING     2257    2257       ATP (BY SIMILARITY).
FT   MOD_RES     2391    2391       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD      77      77       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     401     401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     508     508       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     532     532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     641     641       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     667     667       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     778     778       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     797     797       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     874     874       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     980     980       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1257    1257       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1344    1344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1382    1382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1577    1577       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1587    1587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1665    1665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1752    1752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1776    1776       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1824    1824       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1908    1908       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1966    1966       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    2088    2088       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2594 AA;  289130 MW;  77D8A356CBAD0BBD CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 2594;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     866 AIVSSCV 872
           |||||||
Db    2150 AIVSSCV 2156


RESULT 58
APC_MOUSE
ID   APC_MOUSE      STANDARD;      PRT;  2845 AA.
AC   Q61315; Q62044;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Adenomatous polyposis coli protein (APC protein) (mAPC).
GN   APC.
```

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RC   STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
RX   MEDLINE=92263101; PubMed=1350108;
RA   Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA   Luongo C., Gould K.A., Dove W.F.;
RT   "Multiple intestinal neoplasia caused by a mutation in the murine
RT   homolog of the APC gene.";
RL   Science 256:668-670(1992).
RN   [2]
RP   ERRATUM.
RA   Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA   Luongo C., Gould K.A., Dove W.F.;
RL   Science 256:1114-1114(1992).
RN   [3]
RP   SEQUENCE OF 1-45 FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RA   Dicker F., Lambertz S., Reitmair A., Ballhausen W.G.;
RT   "The murine APC gene: alternative splicing of 5' untranslated
RT   region segments.";
RL   Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=94061824; PubMed=8242607;
RA   Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RT   "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL   Cancer Res. 53:5589-5591(1993).
CC   -!- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC       CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC       ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: FORMS HOMOOLIGOMERS AND ASSOCIATES WITH CATENINS (BY
CC       SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
CC       LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
CC   -!- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M88127; AAB59632.1; -.
DR   EMBL; U02937; AAA03443.1; -.
DR   HSSP; Q02248; 3BCT.
DR   MGD; MGI:88039; Apc.
DR   InterPro; IPR000225; Armadillo.
DR   Pfam; PF00514; Armadillo_seg; 6.
DR   SMART; SM00185; ARM; 5.
DR   PROSITE; PS50176; ARM_REPEAT; 1.
KW   Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;
KW   Coiled coil.
FT   DOMAIN       1     61       COILED COIL (POTENTIAL).
FT   DOMAIN     125    245       COILED COIL (POTENTIAL).
FT   DOMAIN       1    728       LEU-RICH.
FT   REPEAT     451    493       ARM 1.
FT   REPEAT     503    545       ARM 2.
FT   REPEAT     546    589       ARM 3.
FT   REPEAT     590    636       ARM 4.
FT   REPEAT     637    681       ARM 5.
FT   REPEAT     682    723       ARM 6.
FT   REPEAT     724    765       ARM 7.
FT   DOMAIN     739   2834       SER-RICH.
FT   DOMAIN    1130   1156       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN    1556   1575       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN    1864   1891       HIGHLY CHARGED.
FT   VARSPLIC   243    276       MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARSPLIC   310    410       MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT   VARIANT    120    120       T -> A (IN STRAIN CAST/EI).
FT   VARIANT    493    493       V -> I (IN STRAIN CAST/EI).
FT   VARIANT    797    797       Y -> F (IN STRAIN CAST/EI).
FT   VARIANT   1330   1330       A -> T (IN STRAIN CAST/EI).
FT   VARIANT   1618   1618       A -> S (IN STRAIN CAST/EI).
FT   VARIANT   2294   2294       G -> A (IN STRAIN CAST/EI).
FT   VARIANT   2496   2496       H -> Q (IN STRAIN CAST/EI).
FT   VARIANT   2523   2523       T -> A (IN STRAIN CAST/EI).
FT   VARIANT   2813   2813       T -> S (IN STRAIN CAST/EI).
SQ   SEQUENCE   2845 AA;  311086 MW;  145CA73CF570A499 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 2845;
  Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    317 SEKGSSS 323
          |||||||
Db   2687 SEKGSSS 2693


RESULT 59
SACS_HUMAN
ID   SACS_HUMAN     STANDARD;      PRT;  3829 AA.
AC   Q9NZJ4; O94835;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sacsin.
GN   SACS OR KIAA0730.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT ALA-2619.
RX   MEDLINE=20120709; PubMed=10655055;
RA   Engert J.C., Berube P., Mercier J., Dore C., Lepage P., Ge B.,
RA   Bouchard J.-P., Mathieu J., Melancon S.B., Schalling M., Lander E.S.,
RA   Morgan K., Hudson T.J., Richter A.;
RT   "ARSACS, a spastic ataxia common in northeastern Quebec, is caused by
RT   mutations in a new gene encoding an 11.5-kb ORF.";
RL   Nat. Genet. 24:120-125(2000).
RN   [2]
RP   SEQUENCE OF 2826-3829 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99087487; PubMed=9872452;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XI.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 5:277-286(1998).
CC   -!- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE CENTRAL NERVOUS
CC       SYSTEM. ALSO FOUND IN SKELETAL MUSCLE AND AT LOW LEVELS IN
CC       PANCREAS.
CC   -!- DISEASE: DEFECTS IN SACS ARE THE CAUSE OF AUTOSOMAL RECESSIVE
CC       SPASTIC ATAXIA OF CHARLEVOIX-SAGUENAY (ARSACS OR SACS). ARSACS IS
CC       AN EARLY ONSET NEURODEGENERATIVE DISEASE WITH HIGH PREVALENCE IN
CC       THE CHARLEVOIX-SAGUENAY-LAC-SAINT-JEAN REGION OF QUEBEC. IT IS
CC       CHARACTERIZED BY ABSENT SENSORY-NERVE CONDUCTION, REDUCED MOTOR-
CC       NERVE VELOCITY AND HYPERMYELINATION OF RETINAL-NERVE FIBERS.
CC   -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF193556; AAF31262.1; -.
DR   EMBL; AB018273; BAA34450.1; -.
DR   Genew; HGNC:10519; SACS.
DR   MIM; 604490; -.
DR   MIM; 270550; -.
DR   InterPro; IPR001623; DnaJ_N.
DR   PROSITE; PS00636; DNAJ_1; FALSE_NEG.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Chaperone; Polymorphism.
FT   DOMAIN      3556   3643       J-DOMAIN.
FT   VARIANT     2619   2619       V -> A.
FT                                 /FTId=VAR_010296.
SQ   SEQUENCE   3829 AA;  436972 MW;  7AE990311E1E3E91 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 3829;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     652 SLCYNDC 658
           |||||||
Db    1715 SLCYNDC 1721


RESULT 60
SACS_MOUSE
ID   SACS_MOUSE      STANDARD;      PRT;  3830 AA.
AC   Q9JLC8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sacsin.
GN   SACS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20120709; PubMed=10655055;
RA   Engert J.C., Berube P., Mercier J., Dore C., Lepage P., Ge B.,
RA   Bouchard J.-P., Mathieu J., Melancon S.B., Schalling M., Lander E.S.,
RA   Morgan K., Hudson T.J., Richter A.;
RT   "ARSACS, a spastic ataxia common in northeastern Quebec, is caused by
RT   mutations in a new gene encoding an 11.5-kb ORF.";
RL   Nat. Genet. 24:120-125(2000).
CC   -!- FUNCTION: MAY FUNCTION IN CHAPERONE-MEDIATED PROTEIN FOLDING.
CC   -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF193557; AAF31263.1; -.
DR   MGD; MGI:1354724; Sacs.
DR   InterPro; IPR001623; DnaJ_N.
DR   PROSITE; PS00636; DNAJ_1; FALSE_NEG.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Chaperone.
FT   DOMAIN      3557   3644       J-DOMAIN.
SQ   SEQUENCE   3830 AA;  436750 MW;  B3F9DA5A3B0C3EF1 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 3830;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     652 SLCYNDC 658
           |||||||
Db    1715 SLCYNDC 1721


RESULT 61
DYHC_RAT
ID   DYHC_RAT       STANDARD;      PRT;  4644 AA.
AC   P38650; Q63178;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Dynein heavy chain, cytosolic (DYHC) (Cytoplasmic dynein heavy chain)
DE   (MAP 1C).
GN   DNCHC1 OR DNCH1 OR DNEC1 OR MAP1C.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=93376715; PubMed=7690137;
RA   Zhang Z., Tanaka Y., Nonaka S., Aizawa H., Kawasaki H., Nakata T.,
RA   Hirokawa N.;
RT   "The primary structure of rat brain (cytoplasmic) dynein heavy chain,
RT   a cytoplasmic motor enzyme.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:7928-7932(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=93264075; PubMed=7684232;
RA   Mikami A., Paschal B.M., Mazumdar M., Vallee R.B.;
RT   "Molecular cloning of the retrograde transport motor cytoplasmic
RT   dynein (MAP 1C).";
RL   Neuron 10:787-796(1993).
CC   -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC       MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC       ORGANELLES ALONG MICROTUBULES.
CC   -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC       INTERMEDIATE AND LIGHT CHAINS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D13896; BAA02996.1; -.
DR   EMBL; L08505; AAA41103.1; -.
DR   PIR; A38905; A38905.
DR   InterPro; IPR004273; Dynein_heavy.
DR   Pfam; PF03028; Dynein_heavy; 1.
KW   Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
FT   DOMAIN        48     69       COILED COIL (POTENTIAL).
FT   DOMAIN       179    200       COILED COIL (POTENTIAL).
FT   DOMAIN       453    476       COILED COIL (POTENTIAL).
FT   DOMAIN       541    564       COILED COIL (POTENTIAL).
FT   DOMAIN      1169   1201       COILED COIL (POTENTIAL).
FT   DOMAIN      1229   1250       COILED COIL (POTENTIAL).
FT   DOMAIN      1355   1371       COILED COIL (POTENTIAL).
FT   DOMAIN      2012   2040       MICROTUBULE-BINDING (POTENTIAL).
FT   DOMAIN      3187   3273       COILED COIL (POTENTIAL).
FT   DOMAIN      3394   3498       COILED COIL (POTENTIAL).
FT   DOMAIN      3735   3798       COILED COIL (POTENTIAL).
FT   NP_BIND     1904   1911       ATP (POTENTIAL).
FT   NP_BIND     2222   2229       ATP (POTENTIAL).
FT   NP_BIND     2593   2600       ATP (POTENTIAL).
FT   NP_BIND     2935   2942       ATP (POTENTIAL).
```

```
FT   CONFLICT   1024   1025       SR -> MP (IN REF. 2).
FT   CONFLICT   1772   1772       N -> D (IN REF. 2).
FT   CONFLICT   2098   2098       P -> A (IN REF. 2).
FT   CONFLICT   2139   2139       F -> V (IN REF. 2).
FT   CONFLICT   2175   2175       D -> A (IN REF. 2).
FT   CONFLICT   2185   2185       K -> Q (IN REF. 2).
FT   CONFLICT   2366   2366       L -> V (IN REF. 2).
FT   CONFLICT   2382   2382       T -> S (IN REF. 2).
FT   CONFLICT   2463   2463       G -> A (IN REF. 2).
FT   CONFLICT   3219   3219       A -> D (IN REF. 2).
FT   CONFLICT   4131   4131       R -> K (IN REF. 2).
FT   CONFLICT   4366   4366       F -> S (IN REF. 2).
FT   CONFLICT   4511   4511       A -> G (IN REF. 2).
SQ   SEQUENCE   4644 AA;  532240 MW;  8C6ABDBEDF875D82 CRC64;

  Query Match             0.7%;  Score 7;  DB 1;  Length 4644;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 SARVRGR 17
           |||||||
Db     720 SARVRGR 726


RESULT 62
PSBL_ARATH
ID   PSBL_ARATH     STANDARD;      PRT;    37 AA.
AC   P29301;
DT   01-DEC-1992 (Rel. 24, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Photosystem II reaction center L protein (PSII 5 kDa protein).
GN   PSBL.
OS   Arabidopsis thaliana (Mouse-ear cress), and
OS   Antirrhinum majus (Garden snapdragon).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702, 4151;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.majus; STRAIN=cv. Sippe 50; TISSUE=Leaf;
RX   MEDLINE=92191997; PubMed=1547774;
RA   Kudla J., Igloi G.L., Metzlaff M., Hagenmann H., Koessel H.;
RT   "RNA editing in tobacco chloroplasts leads to the formation of a
RT   translatable psbL mRNA by a C to U substitution within the initiation
RT   codon.";
RL   EMBO J. 11:1099-1103(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.thaliana; STRAIN=cv. Columbia;
RX   MEDLINE=20039611; PubMed=10574454;
RA   Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
RT   "Complete structure of the chloroplast genome of Arabidopsis
RT   thaliana.";
RL   DNA Res. 6:283-290(1999).
CC   -!- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
CC       B559.
CC   -!- SIMILARITY: BELONGS TO THE PSBL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63206; CAA44889.1; -.
DR   EMBL; AP000423; BAA84400.1; -.
DR   PIR; S19786; F2SKL.
DR   InterPro; IPR003372; PSII_PsbL.
DR   Pfam; PF02419; PsbL; 1.
KW   Photosynthesis; Photosystem II; Chloroplast.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   37 AA;  4339 MW;  CC537AEC4C630A84 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     232 SVELNR 237
           ||||||
Db       9 SVELNR 14


RESULT 63
TX21_SELHU
ID   TX21_SELHU     STANDARD;      PRT;    37 AA.
AC   P82959;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Huwentoxin-II isoform 1 (HwTx-II).
OS   Selenocosmia huwena (Chinese bird spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Mygalomorphae; Theraphosidae; Ornithoctonus.
OX   NCBI_TaxID=29017;
RN   [1]
RP   SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RC   TISSUE=Venom;
RX   MEDLINE=99351628; PubMed=10424342;
RA   Shu Q., Liang S.-P.;
RT   "Purification and characterization of huwentoxin-II, a neurotoxic
RT   peptide from the venom of the Chinese bird spider Selenocosmia
RT   huwena.";
RL   J. Pept. Res. 53:486-491(1999).
CC   -!- FUNCTION: LETHAL NEUROTOXIN. BINDS TO THE NICOTINIC ACETYLCHOLINE
CC       RECEPTOR. BLOCKS NEUROMUSCULAR TRANSMISSION. ACTS COOPERATIVELY TO
CC       POTENTIATE THE ACTIVITY OF HUWENTOXIN-I.
CC   -!- SUBUNIT: ISOFORM 1 AND ISOFORM 2 MAY DIMERIZE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- MASS SPECTROMETRY: MW=4290.3; METHOD=MALDI.
CC   -!- SIMILARITY: BELONGS TO THE HUWENTOXIN-II FAMILY.
KW   Toxin; Neurotoxin; Postsynaptic neurotoxin.
FT   DISULFID      4     18
FT   DISULFID      8     29
FT   DISULFID     23     34
SQ   SEQUENCE   37 AA;  4290 MW;  E0F9F84321AD6382 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      90 ECSFSC 95
           ||||||
Db       3 ECSFSC 8


RESULT 64
TX22_SELHU
ID   TX22_SELHU     STANDARD;      PRT;    37 AA.
AC   P82960;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Huwentoxin-II isoform 2 (HwTx-II).
OS   Selenocosmia huwena (Chinese bird spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Mygalomorphae; Theraphosidae; Ornithoctonus.
OX   NCBI_TaxID=29017;
RN   [1]
RP   SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RC   TISSUE=Venom;
```

```
RX   MEDLINE=99351628; PubMed=10424342;
RA   Shu Q., Liang S.-P.;
RT   "Purification and characterization of huwentoxin-II, a neurotoxic
RT   peptide from the venom of the Chinese bird spider Selenocosmia
RT   huwena.";
RL   J. Pept. Res. 53:486-491(1999).
CC   -!- FUNCTION: LETHAL NEUROTOXIN. BINDS TO THE NICOTINIC ACETYLCHOLINE
CC       RECEPTOR. BLOCKS NEUROMUSCULAR TRANSMISSION. ACTS COOPERATIVELY TO
CC       POTENTIATE THE ACTIVITY OF HUWENTOXIN-I.
CC   -!- SUBUNIT: ISOFORM 1 AND ISOFORM 2 MAY DIMERIZE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- MASS SPECTROMETRY: MW=4305.2; METHOD=MALDI.
CC   -!- SIMILARITY: BELONGS TO THE HUWENTOXIN-II FAMILY.
KW   Toxin; Neurotoxin; Postsynaptic neurotoxin.
FT   DISULFID      4     18
FT   DISULFID      8     29
FT   DISULFID     23     34
SQ   SEQUENCE   37 AA;  4305 MW;  E0F9F85AA0356382 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 37;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      90 ECSFSC 95
           ||||||
Db       3 ECSFSC 8


RESULT 65
PSBL_CHLVU
ID   PSBL_CHLVU     STANDARD;      PRT;    38 AA.
AC   P56339;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Photosystem II reaction center L protein (PSII 5 kDa protein).
GN   PSBL.
OS   Chlorella vulgaris.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Chlorella.
OX   NCBI_TaxID=3077;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IAM C-27 / Tamiya;
RX   MEDLINE=97303241; PubMed=9159184;
RA   Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA   Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA   Inamura A., Yoshinaga K., Sugiura M.;
RT   "Complete nucleotide sequence of the chloroplast genome from the
RT   green alga Chlorella vulgaris: the existence of genes possibly
RT   involved in chloroplast division.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC   -!- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
CC       B559.
CC   -!- SIMILARITY: BELONGS TO THE PSBL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB001684; BAA57901.1; -.
DR   InterPro; IPR003372; PSII_PsbL.
DR   Pfam; PF02419; PsbL; 1.
KW   Photosynthesis; Photosystem II; Chloroplast.
SQ   SEQUENCE   38 AA;  4388 MW;  DF737195D0DEA61D CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     232 SVELNR 237
           ||||||
Db      10 SVELNR 15


RESULT 66
PSBL_MARPO
ID   PSBL_MARPO     STANDARD;      PRT;    38 AA.
AC   P12165;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Photosystem II reaction center L protein (PSII 5 kDa protein).
GN   PSBL.
OS   Marchantia polymorpha (Liverwort).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC   Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89068687; PubMed=3199436;
RA   Fukuzawa H., Kohchi T., Sano T., Shirai H., Umesono K., Inokuchi H.,
RA   Ozeki H., Ohyama K.;
RT   "Structure and organization of Marchantia polymorpha chloroplast
RT   genome. III. Gene organization of the large single copy region from
RT   rbcL to trnI(CAU).";
RL   J. Mol. Biol. 203:333-351(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA   Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA   Ozeki H.;
RT   "Chloroplast gene organization deduced from complete sequence of
RT   liverwort Marchantia polymorpha chloroplast DNA.";
RL   Nature 322:572-574(1986).
CC   -!- FUNCTION: NOT KNOWN. POSSIBLY A PERIPHERAL COMPONENT OF CYTOCHROME
CC       B559.
CC   -!- SIMILARITY: BELONGS TO THE PSBL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04465; CAA28099.1; -.
DR   PIR; A05048; A05048.
DR   PIR; S01538; S01538.
DR   InterPro; IPR003372; PSII_PsbL.
DR   Pfam; PF02419; PsbL; 1.
KW   Photosynthesis; Photosystem II; Chloroplast.
SQ   SEQUENCE   38 AA;  4479 MW;  DF73781163D39D6D CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 76;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     232 SVELNR 237
           ||||||
Db      10 SVELNR 15


RESULT 67
PIV6_ADEB2
ID   PIV6_ADEB2     STANDARD;      PRT;    43 AA.
AC   Q96627;
```

```
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Minor capsid protein VI precursor (Fragment).
GN   PVI.
OS   Bovine adenovirus type 2 (Mastadenovirus bos2).
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=114429;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rusvai M., Harrach B., Banrevi A., Evans P., Benko M.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MINOR CAPSID PROTEIN THAT MAY ACT AS A LINK BETWEEN THE
CC       EXTERNAL CAPSID AND THE INTERNAL DNA-PROTEIN CORE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U44123; AAB16759.1; -.
DR   InterPro; IPR004243; McpVI.
DR   Pfam; PF02993; MCPVI; 1.
FT   PROPEP        1     33       BY SIMILARITY.
FT   CHAIN        34    >43       MINOR CAPSID PROTEIN VI.
FT   NON_TER      43     43
SQ   SEQUENCE   43 AA;  4584 MW;  10F78E9678070306 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 43;
  Best Local Similarity   100.0%;  Pred. No. 85;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     672 NFSALA 677
           ||||||
Db       5 NFSALA 10


RESULT 68
ATP8_YARLI
ID   ATP8_YARLI     STANDARD;      PRT;    48 AA.
AC   Q36257;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN   ATP8.
OS   Yarrowia lipolytica (Candida lipolytica).
OG   Mitochondrion.
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Dipodascaceae; Yarrowia.
OX   NCBI_TaxID=4952;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 44601;
RA   Matsuoka M., Matsubara M., Inoue J., Kakehi M., Imanaka T.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC       (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L15359; AAA78260.1; -.
KW   Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT   TRANSMEM     13     32       POTENTIAL.
SQ   SEQUENCE   48 AA;  5832 MW;  6076BA02936A0BF7 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 94;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     133 GFASLS 138
           ||||||
Db      16 GFASLS 21


RESULT 69
ATP8_PARLI
ID   ATP8_PARLI     STANDARD;      PRT;    54 AA.
AC   P12697;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).
GN   MTATP8 OR ATP8.
OS   Paracentrotus lividus (Common sea urchin).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
OC   Paracentrotus.
OX   NCBI_TaxID=7656;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89291831; PubMed=2544576;
RA   Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.;
RT   "The complete nucleotide sequence, gene organization, and genetic
RT   code of the mitochondrial genome of Paracentrotus lividus.";
RL   J. Biol. Chem. 264:10965-10975(1989).
CC   -!- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC       (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
CC       H(+)(Out).
CC   -!- SUBCELLULAR LOCATION: Membrane-bound.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04815; AAA68138.1; -.
DR   PIR; F34284; F34284.
DR   InterPro; IPR001421; ATPase8_mit.
DR   Pfam; PF00895; ATP-synt_8; 1.
KW   Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
SQ   SEQUENCE   54 AA;  6452 MW;  41E52C808607EA46 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     919 TAILLT 924
           ||||||
Db      18 TAILLT 23


RESULT 70
Y737_ARCFU
ID   Y737_ARCFU     STANDARD;      PRT;    67 AA.
AC   O29521;
```

```
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AF0737.
GN   AF0737.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001054; AAB90514.1; -.
DR   TIGR; AF0737; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   67 AA;  7296 MW;  6F103AABBBE56630 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 67;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     787 LESLGI 792
           ||||||
Db      28 LESLGI 33


RESULT 71
RUXF_SCHPO
ID   RUXF_SCHPO     STANDARD;      PRT;    78 AA.
AC   O59734;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable small nuclear ribonucleoprotein F (snRNP-F) (Sm protein F)
DE   (Sm-F) (SmF).
GN   SMF1 OR SPBC3E7.14 OR SPBC4F6.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: PROBABLE COMMON SM PROTEIN, IS FOUND IN U1 AND U2
CC       SNRNPS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL023534; CAA19017.1; -.
DR   EMBL; AL031534; CAA20721.1; -.
DR   InterPro; IPR001163; snRNP_Sm.
DR   Pfam; PF01423; Sm; 1.
KW   Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
KW   RNA-binding.
SQ   SEQUENCE   78 AA;  8660 MW;  005687FA91F013CA CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 78;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     257 KPVLVR 262
           ||||||
Db      18 KPVLVR 23


RESULT 72
NSGX_HUMAN
ID   NSGX_HUMAN     STANDARD;      PRT;    79 AA.
AC   Q9UH64;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Susceptibility protein NSG-x.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Nasopharyngeal carcinoma;
RA   Jun Q., Yang J.B., Li G.Y.;
RT   "A new nasopharyngeal carcinoma associated gene on 9p21-22.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Burian D.M., Mitchell N., Roe B.A.;
```

```
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF211119; AAF22974.1; -.
DR   EMBL; AC000048; -; NOT_ANNOTATED_CDS.
SQ   SEQUENCE   79 AA;  8851 MW;  EC94C8BC34FD5F8C CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 79;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 RIPRLW 26
           ||||||
Db      64 RIPRLW 69


RESULT 73
YVKA_VACCC
ID   YVKA_VACCC     STANDARD;      PRT;    81 AA.
AC   P20569;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 9.4 kDa protein.
GN   K ORF A.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48007.1; -.
DR   PIR; G42505; G42505.
KW   Hypothetical protein.
SQ   SEQUENCE   81 AA;  9386 MW;  42CE3D0531FEA9C4 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 81;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     571 KIYSIN 576
           ||||||
Db      73 KIYSIN 78


RESULT 74
Y67_BPT3
ID   Y67_BPT3       STANDARD;      PRT;    83 AA.
AC   P20330;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   Hypothetical gene 6.7 protein.
GN   6.7.
OS   Bacteriophage T3.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC   T7-like viruses.
OX   NCBI_TaxID=10759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Luria;
RX   MEDLINE=90133923; PubMed=2614843;
RA   Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT   "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL   J. Mol. Biol. 210:687-701(1989).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17255; CAA35150.1; -.
DR   PIR; S07519; S07519.
KW   Hypothetical protein.
SQ   SEQUENCE   83 AA;  8848 MW;  C4953F5806081581 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 83;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     143 LDDSAA 148
           ||||||
Db      64 LDDSAA 69


RESULT 75
Y425_TREPA
ID   Y425_TREPA     STANDARD;      PRT;    86 AA.
AC   O83440;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein TP0425.
GN   TP0425.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE001220; AAC65419.1; -.
DR   TIGR; TP0425; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   86 AA;  9737 MW;  728CDB709B7132DB CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 86;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      62 DSTGSR 67
           ||||||
Db       8 DSTGSR 13


RESULT 76
FXY4_MOUSE
ID   FXY4_MOUSE     STANDARD;      PRT;    88 AA.
AC   Q9D2W0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   FXYD domain-containing ion transport regulator 4 precursor (Channel
DE   inducing factor) (CHIF).
GN   FXYD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Garty H.;
RT   "Genomic sequence of mouse CHIF (FXYD4).";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE FXYD FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF362729; AAK51508.1; -.
DR   EMBL; AK018728; BAB31372.1; -.
DR   MGD; MGI:1889005; Fxyd4.
DR   InterPro; IPR000272; ATP1G1_PLM_MAT8.
DR   Pfam; PF02038; ATP1G1_PLM_MAT8; 1.
DR   PROSITE; PS01310; FXYD; 1.
KW   Transmembrane; Signal; Ionic channel; Ion transport.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21     88       FXYD DOMAIN-CONTAINING ION TRANSPORT
FT                                REGULATOR 4.
FT   DOMAIN       21     38       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     39     59       POTENTIAL.
FT   DOMAIN       60     88       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   88 AA;  9269 MW;  7EB0140941CFE926 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 88;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     592 ALEASD 597
           ||||||
Db      17 ALEASD 22


RESULT 77
XHLA_BACSU
ID   XHLA_BACSU     STANDARD;      PRT;    89 AA.
AC   P39798;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Protein xhlA.
GN   XHLA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / SO113;
RA   Krogh S., Joergensen S.T., Diderichsen B., Devine K.M.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95005440; PubMed=7921239;
RA   Longchamp P.F., Mauel C., Karamata D.;
RT   "Lytic enzymes associated with defective prophages of Bacillus
RT   subtilis: sequencing and characterization of the region comprising
RT   the N-acetylmuramoyl-L-alanine amidase gene of prophage PBSX.";
RL   Microbiology 140:1855-1867(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98215178; PubMed=9555893;
RA   Krogh S., Jorgensen S.T., Devine K.M.;
RT   "Lysis genes of the Bacillus subtilis defective prophage PBSX.";
RL   J. Bacteriol. 180:2110-2117(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
```

```
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: ASSOCIATED WITH CELL LYSIS UPON INDUCTION OF PBSX.
CC   -!- SIMILARITY: STRONG, TO B.LICHENIFORMIS XPAF1 AND XPAL1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z36941; CAA85401.1; -.
DR   EMBL; L25924; AAA22643.1; -.
DR   EMBL; Z70177; CAA94047.1; -.
DR   EMBL; Z99110; CAB13136.1; -.
DR   PIR; S47314; S47314.
DR   SubtiList; BG10960; xhlA.
KW   Transmembrane; Complete proteome.
FT   TRANSMEM     63     83       POTENTIAL.
SQ   SEQUENCE   89 AA;  9997 MW;  E8BB53B7A3B81AEA CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 89;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     541 IEENTT 546
           ||||||
Db      52 IEENTT 57


RESULT 78
YAL1_BACLI
ID   YAL1_BACLI     STANDARD;      PRT;    89 AA.
AC   Q99164;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 10.1 kDa protein in ORF3 5'region.
GN   XPAL1.
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC14;
RX   MEDLINE=91237328; PubMed=2033382;
RA   Lee J.W.K., Edwards C.W., Hulett F.M.;
RT   "Identification of four unique clones encoding 10 kDa proteins from
RT   Bacillus that cause phenotypic complementation of a phoA mutant
RT   strain of Escherichia coli.";
RL   J. Gen. Microbiol. 137:667-677(1991).
CC   -!- SIMILARITY: TO B.LICHENIFORMIS XPAF1 AND TO B.SUBTILIS XHLA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63942; AAA22886.1; -.
DR   PIR; B49754; B49754.
KW   Hypothetical protein.
SQ   SEQUENCE   89 AA;  10071 MW;  6686426FB93E6ABE CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 89;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     541 IEENTT 546
           ||||||
Db      52 IEENTT 57


RESULT 79
ES6B_MYCTU
ID   ES6B_MYCTU     STANDARD;      PRT;    95 AA.
AC   O05454;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative ESAT-6 like protein 11.
GN   RV3890C OR MT4005 OR MTCY15F10.22.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z94121; CAB08084.1; -.
DR   EMBL; AE007192; AAK48372.1; -.
DR   TIGR; MT4005; -.
```

```
DR   TubercuList; Rv3890c; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   95 AA;  9920 MW;  E554BBFAE55C56F5 CRC64;

  Query Match           0.6%;  Score 6;  DB 1;  Length 95;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     595 ASDVGS 600
           ||||||
Db      16 ASDVGS 21


RESULT 80
XPA_CRIGR
ID   XPA_CRIGR      STANDARD;      PRT;    97 AA.
AC   Q64029;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-repair protein complementing XP-A cells homolog (Xeroderma
DE   pigmentosum group A complementing protein homolog) (Fragment).
GN   XPA OR XPAC.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=95064305; PubMed=7974007;
RA   Cleaver J.E., McDowell M., Jones C., Wood R., Karentz D.;
RT   "Mutation and expression of the XPA gene in revertants and hybrids of
RT   a Xeroderma pigmentosum cell line.";
RL   Somat. Cell Mol. Genet. 20:327-337(1994).
CC   -!- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. INITIATES REPAIR BY
CC       BINDING TO DAMAGED SITES WITH VARIOUS AFFINITIES, DEPENDING ON THE
CC       PHOTOPRODUCT AND THE TRANSCRIPTIONAL STATE OF THE REGION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE XPA FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; S74024; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR000465; XPA_protein.
DR   Pfam; PF01286; XPA; 1.
DR   TIGRFAMs; TIGR00598; rad14; 1.
DR   PROSITE; PS00752; XPA_1; PARTIAL.
DR   PROSITE; PS00753; XPA_2; 1.
KW   DNA repair; DNA-binding; Zinc-finger; Nuclear protein.
FT   NON_TER       1      1
FT   ZN_FING      <1     15
FT   NON_TER      97     97
SQ   SEQUENCE   97 AA;  11519 MW;  B55360D5C5C24EBE CRC64;

  Query Match           0.6%;  Score 6;  DB 1;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     950 LKDCDL 955
           ||||||
Db      36 LKDCDL 41


RESULT 81
YAF1_BACLI
ID   YAF1_BACLI     STANDARD;      PRT;    97 AA.
AC   P37135;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 11.0 kDa protein in CWLL 5'region.
GN   XPAF1.
OS   Bacillus licheniformis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1402;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FD0120;
RX   MEDLINE=94067020; PubMed=7902527;
RA   Oda Y., Nakayama R., Kuroda A., Sekiguchi J.;
RT   "Molecular cloning, sequence analysis, and characterization of a new
RT   cell wall hydrolase, CwlL, of Bacillus licheniformis.";
RL   Mol. Gen. Genet. 241:380-388(1993).
CC   -!- SIMILARITY: TO B.LICHENIFORMIS XPAL1 AND TO B.SUBTILIS XHLA.
CC   -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D13377; BAA02644.1; ALT_INIT.
DR   EMBL; D13377; BAA02645.1; ALT_INIT.
DR   PIR; S39914; S39914.
KW   Hypothetical protein.
SQ   SEQUENCE   97 AA;  11083 MW;  7811F1E680DB7369 CRC64;

  Query Match           0.6%;  Score 6;  DB 1;  Length 97;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     541 IEENTT 546
           ||||||
Db      60 IEENTT 65


RESULT 82
RK23_ASTLO
ID   RK23_ASTLO     STANDARD;      PRT;    98 AA.
AC   P34771;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chloroplast 50S ribosomal protein L23.
GN   RPL23.
OS   Astasia longa (Euglenophycean alga).
OG   Chloroplast.
OC   Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.
OX   NCBI_TaxID=3037;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 1204-17a;
RX   MEDLINE=95062725; PubMed=7972503;
RA   Gockel G., Baier S., Hachtel W.;
RT   "Plastid ribosomal protein genes from the nonphotosynthetic
RT   flagellate Astasia longa.";
RL   Plant Physiol. 105:1443-1444(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 1204-17a;
RX   MEDLINE=21080522; PubMed=11212895;
RA   Gockel G., Hachtel W.;
RT   "Complete gene map of the plastid genome of the nonphotosynthetic
RT   euglenoid flagellate Astasia longa.";
RL   Protist 151:347-351(2000).
```

```
CC   -!- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ294725; CAC24593.1; -.
DR   PIR; S38604; S38604.
DR   InterPro; IPR001014; Ribosomal_L23.
DR   Pfam; PF00276; Ribosomal_L23; 1.
DR   ProDom; PD001141; Ribosomal_L23; 1.
DR   PROSITE; PS00050; RIBOSOMAL_L23; 1.
KW   Ribosomal protein; Chloroplast; rRNA-binding.
SQ   SEQUENCE   98 AA;  11944 MW;  7572EB84DA96D347 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 98;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     571 KIYSIN 576
           ||||||
Db      54 KIYSIN 59


RESULT 83
YNI6_YEAST
ID   YNI6_YEAST     STANDARD;      PRT;   102 AA.
AC   P48232;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 11.5 kDa protein in TOP2-MKT1 intergenic region.
GN   YNL086W OR N2254.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96310628; PubMed=8740422;
RA   Soler-Mira A., Saiz J.E., Ballesta J.P.G., Remacha M.;
RT   "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae
RT   chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five
RT   new open reading frames.";
RL   Yeast 12:485-491(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X89016; CAA61424.1; -.
DR   EMBL; Z71362; CAA95962.1; -.
DR   SGD; S0005030; YNL086W.
KW   Hypothetical protein.
SQ   SEQUENCE   102 AA;  11461 MW;  33200D2F0F729DA1 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 102;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 LSARVR 15
           ||||||
Db      79 LSARVR 84


RESULT 84
GLN1_METTM
ID   GLN1_METTM     STANDARD;      PRT;   105 AA.
AC   Q50786;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Nitrogen fixation nifHD region glnB-like protein 1.
GN   GLNBA.
OS   Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=79929;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96163477; PubMed=8575452;
RA   Hochheimer A., Schmitz R.A., Thauer R.K., Hedderich R.;
RT   "The tungsten formylmethanofuran dehydrogenase from Methanobacterium
RT   thermoautotrophicum contains sequence motifs characteristic for
RT   enzymes containing molybdopterin dinucleotide.";
RL   Eur. J. Biochem. 234:910-920(1995).
CC   -!- FUNCTION: COULD BE INVOLVED IN THE REGULATION OF NITROGEN
CC       FIXATION.
CC   -!- SIMILARITY: BELONGS TO THE P(II) PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X87971; CAA61217.1; -.
DR   HSSP; P05826; 2PII.
DR   InterPro; IPR002187; PII_glnB.
DR   Pfam; PF00543; P-II; 1.
DR   PRINTS; PR00340; PIIGLNB.
DR   ProDom; PD001194; PII_glnB; 1.
DR   PROSITE; PS00638; PII_GLNB_CTER; 1.
KW   Transcription regulation; Nitrogen fixation.
SQ   SEQUENCE   105 AA;  11635 MW;  35DD1AD0994E1FC3 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     121 GIRFDE 126
           ||||||
Db      41 GIRFDE 46


RESULT 85
HIS3_LISIN
ID   HIS3_LISIN     STANDARD;      PRT;   105 AA.
AC   Q92E89;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH).
GN   HISI OR LIN0571.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / Serovar 6a;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
```

```
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
CC   -!- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-
CC       phosphoribosyl)-5-[(5-
CC       phosphoribosylamino)methylideneamino]imidazole-4-carboxamide.
CC   -!- PATHWAY: Histidine biosynthesis; third step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PRA-CH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL596165; CAC95803.1; -.
DR   ListiList; LIN00571; -.
DR   InterPro; IPR002496; PRA-CH.
DR   Pfam; PF01502; PRA-CH; 1.
DR   ProDom; PD002610; PRA-CH; 1.
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
SQ   SEQUENCE   105 AA;  12109 MW;  BEB644851CF55FA3 CRC64;

  Query Match              0.6%;  Score 6;  DB 1;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     302 NKGETS 307
           ||||||
Db      55 NKGETS 60


RESULT 86
THI1_SYNY3
ID   THI1_SYNY3     STANDARD;      PRT;   105 AA.
AC   P52232;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thioredoxin-like protein slr0233.
GN   SLR0233.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D64000; BAA10238.1; -.
DR   HSSP; P80579; 1QUW.
DR   InterPro; IPR000063; Thiored.
DR   Pfam; PF00085; thiored; 1.
DR   PRINTS; PR00421; THIOREDOXIN.
DR   TIGRFAMs; TIGR01068; thioredoxin; 1.
DR   PROSITE; PS00194; THIOREDOXIN; 1.
KW   Hypothetical protein; Redox-active center; Electron transport;
KW   Complete proteome.
FT   DISULFID     30     33       REDOX-ACTIVE (BY SIMILARITY).
SQ   SEQUENCE   105 AA;  11802 MW;  19958B167EFAAC13 CRC64;

  Query Match              0.6%;  Score 6;  DB 1;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     256 PKPVLV 261
           ||||||
Db      18 PKPVLV 23


RESULT 87
Y793_SYNY3
ID   Y793_SYNY3     STANDARD;      PRT;   108 AA.
AC   Q55939;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein sll0793.
GN   SLL0793.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0060 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D64005; BAA10705.1; -.
DR   InterPro; IPR003844; UPF0060.
DR   Pfam; PF02694; UPF0060; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      7     27       POTENTIAL.
FT   TRANSMEM     32     52       POTENTIAL.
FT   TRANSMEM     64     84       POTENTIAL.
FT   TRANSMEM     86    106       POTENTIAL.
SQ   SEQUENCE   108 AA;  11998 MW;  0D6D83AF5A2358A6 CRC64;

  Query Match              0.6%;  Score 6;  DB 1;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     920 AILLTV 925
           ||||||
Db      40 AILLTV 45


RESULT 88
VIF_HV1SC
```

```
ID   VIF_HV1SC      STANDARD;      PRT;   109 AA.
AC   P05899;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-JUL-1993 (Rel. 26, Last annotation update)
DE   Virion infectivity factor (SOR protein) (Fragment).
GN   VIF.
OS   Human immunodeficiency virus type 1 (SC isolate) (HIV-1).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88219542; PubMed=3369091;
RA   Gurgo C., Guo H.-G., Franchini G., Aldovini A., Collalti E.,
RA   Farrell K., Wong-Staal F., Gallo R.C., Reitz M.S. Jr.;
RT   "Envelope sequences of two new United States HIV-1 isolates.";
RL   Virology 164:531-536(1988).
CC   -!- FUNCTION: DETERMINES VIRUS INFECTIVITY.
CC   -!- MISCELLANEOUS: THE SC ISOLATE WAS TAKEN FROM AN ARC PATIENT IN
CC       1984 IN SOUTHERN CALIFORNIA.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M17450; AAA45059.1; -.
DR   HIV; M17450; VIF$SC.
DR   InterPro; IPR000475; Viral_infect.
DR   Pfam; PF00559; Vif; 1.
DR   ProDom; PD000063; Viral_infect; 1.
KW   AIDS.
FT   NON_TER       1      1
SQ   SEQUENCE   109 AA;  12308 MW;  3D1BB3599F78B727 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     762 LADRLI 767
           ||||||
Db      19 LADRLI 24


RESULT 89
RLA1_DROME
ID   RLA1_DROME     STANDARD;      PRT;   112 AA.
AC   P08570; Q9VPP6;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60S acidic ribosomal protein P1 (RP21C) (Acidic ribosomal protein
DE   RPA2).
GN   RPP2 OR M(2)21C OR RPA2 OR RP21C OR CG4087.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88096510; PubMed=3122177;
RA   Wigboldus J.D.;
RT   "cDNA and deduced amino acid sequence of Drosophila rp21C, another
RT   'A'-type ribosomal protein.";
RL   Nucleic Acids Res. 15:10064-10064(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93273819; PubMed=8501137;
RA   Olson P.F., Salo T., Garrison K., Fessler J.H.;
RT   "Drosophila acidic ribosomal protein rpA2: sequence and
RT   characterization.";
RL   J. Cell. Biochem. 51:353-359(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC       PROTEIN SYNTHESIS.
CC   -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC       SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y00504; CAA68557.1; -.
DR   EMBL; S62170; AAB26902.1; -.
DR   EMBL; AE003589; AAF51499.1; -.
DR   PIR; S00659; R5FF2E.
DR   FlyBase; FBgn0002593; RpP2.
DR   InterPro; IPR001813; 60s_ribosomal.
DR   Pfam; PF00428; 60s_ribosomal; 1.
KW   Ribosomal protein.
FT   CONFLICT      9      9       C -> S (IN REF. 1).
FT   CONFLICT     53     53       G -> A (IN REF. 1).
SQ   SEQUENCE   112 AA;  11513 MW;  2EA9CA3E884A7CCF CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 112;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
```

```
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    624 NTILKA 629
          ||||||
Db     29 NTILKA 34


RESULT 90
V195_FOWPV
ID   V195_FOWPV      STANDARD;      PRT;   113 AA.
AC   Q9J538;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein FPV195.
GN   FPV195.
OS   Fowlpox virus (FPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Avipoxvirus.
OX   NCBI_TaxID=10261;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20193820; PubMed=10729156;
RA   Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT   "The genome of fowlpox virus.";
RL   J. Virol. 74:3815-3831(2000).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A31 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF198100; AAF44539.1; -.
SQ   SEQUENCE   113 AA;  13609 MW;  6973B069675B3328 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    189 VNPEYY 194
          ||||||
Db     77 VNPEYY 82


RESULT 91
HIS3_CLOAB
ID   HIS3_CLOAB      STANDARD;      PRT;   115 AA.
AC   Q97KH7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19) (PRA-CH).
GN   HISI OR CAC0942.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
CC   -!- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-
CC       phosphoribosyl)-5-[(5-
CC       phosphoribosylamino)methylideneamino]imidazole-4-carboxamide.
CC   -!- PATHWAY: Histidine biosynthesis; third step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PRA-CH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE007609; AAK78918.1; -.
DR   InterPro; IPR002496; PRA-CH.
DR   Pfam; PF01502; PRA-CH; 1.
DR   ProDom; PD002610; PRA-CH; 1.
KW   Histidine biosynthesis; Hydrolase; Complete proteome.
SQ   SEQUENCE   115 AA;  13249 MW;  4C735C1F0EF09BBD CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 115;
  Best Local Similarity   100.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    302 NKGETS 307
          ||||||
Db     61 NKGETS 66


RESULT 92
VP3_CAV26
ID   VP3_CAV26      STANDARD;      PRT;   120 AA.
AC   P54095;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Apoptin (VP3).
OS   Chicken anemia virus (USA isolate 26p4) (CAV).
OC   Viruses; ssDNA viruses; Circoviridae; Circovirus.
OX   NCBI_TaxID=73477;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341490; PubMed=1908516;
RA   Claessens J.A.J., Schrier C.C., Mockett A.P.A., Jagt E.H.J.M.,
RA   Sondermeijer P.J.A.;
RT   "Molecular cloning and sequence analysis of the genome of chicken
RT   anaemia agent.";
RL   J. Gen. Virol. 72:2003-2006(1991).
CC   -!- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC       IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10068; -; NOT_ANNOTATED_CDS.
KW   Apoptosis; Nuclear protein.
SQ   SEQUENCE   120 AA;  13061 MW;  42AB3041581A045E CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    700 TLSLCG 705
          ||||||
Db     42 TLSLCG 47
```

```
RESULT 93
SR14_ARATH
ID   SR14_ARATH     STANDARD;      PRT;   121 AA.
AC   O04421; O22839;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Signal recognition particle 14 kDa protein (SRP14).
GN   SRP14 OR AT2G43640 OR F18O19.25.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia; TISSUE=Root;
RA   Bui N., Wolff N., Strub K.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC       IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
CC       MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE ALU PORTION OF THE SRP
CC       RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX
CC       OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).
CC   -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC       OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC       SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE SRP14 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y10116; CAA71202.1; -.
DR   EMBL; AC002333; AAB64042.1; -.
DR   HSSP; P16254; 1914.
DR   InterPro; IPR003210; SRP14.
DR   Pfam; PF02290; SRP14; 1.
KW   Signal recognition particle; RNA-binding.
FT   CONFLICT    116    116       P -> T (IN REF. 2).
SQ   SEQUENCE   121 AA;  13777 MW;  216D2AA83B24E7DD CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     625 TILKAH 630
           ||||||
Db      84 TILKAH 89


RESULT 94
```

```
VP3_CAV82
ID   VP3_CAV82      STANDARD;      PRT;   121 AA.
AC   P54096;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Apoptin (VP3).
OS   Chicken anemia virus (Japanese isolate 82-2) (CAV).
OC   Viruses; ssDNA viruses; Circoviridae; Circovirus.
OX   NCBI_TaxID=73476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95297149; PubMed=7778281;
RA   Kato A., Fujino M., Nakamura T., Ishihama A., Otaki Y.;
RT   "Gene organization of chicken anemia virus.";
RL   Virology 209:480-488(1995).
CC   -!- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC       IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D31965; BAA06733.1; -.
KW   Apoptosis; Nuclear protein.
SQ   SEQUENCE   121 AA;  13233 MW;  C770839C2BD61A3E CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     700 TLSLCG 705
           ||||||
Db      43 TLSLCG 48


RESULT 95
VP3_CAVC1
ID   VP3_CAVC1      STANDARD;      PRT;   121 AA.
AC   Q99152;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Apoptin (VP3).
OS   Chicken anemia virus (German isolate Cuxhaven-1) (CAV).
OC   Viruses; ssDNA viruses; Circoviridae; Circovirus.
OX   NCBI_TaxID=73475;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91237831; PubMed=1851873;
RA   Noteborn M.H.M., de Boer G.F., van Roozelaar D.J., Karreman C.,
RA   Kranenburg O., Vos J.G., Jeurissen S.H.M., Hoeben R.C., Zantema A.,
RA   Koch G., van Ormondt H., van der Eb A.J.;
RT   "Characterization of cloned chicken anemia virus DNA that contains
RT   all elements for the infectious replication cycle.";
RL   J. Virol. 65:3131-3139(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92296898; PubMed=1605740;
RA   Meehan B.M., Todd D., Creelan J.L., Earle J.A.P., Hoey E.M.,
RA   McNulty M.S.;
RT   "Characterization of viral DNAs from cells infected with chicken
RT   anaemia agent: sequence analysis of the cloned replicative form and
RT   transfection capabilities of cloned genome fragments.";
RL   Arch. Virol. 124:301-319(1992).
CC   -!- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC       IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55918; AAA91823.1; -.
DR   EMBL; M81223; AAA42883.1; -.
KW   Apoptosis; Nuclear protein.
FT   CONFLICT     70     70       F -> S (IN REF. 2).
FT   CONFLICT    116    116       K -> R (IN REF. 2).
FT   CONFLICT    118    118       R -> C (IN REF. 2).
SQ   SEQUENCE   121 AA;  13270 MW;  D2AAB3869BC12A3E CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     700 TLSLCG 705
           ||||||
Db      43 TLSLCG 48


RESULT 96
VP3_CAVCI
ID   VP3_CAVCI      STANDARD;      PRT;   121 AA.
AC   P54094;
DT   01-OCT-1996 (Rel. 34, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Apoptin (VP3).
OS   Chicken anemia virus (USA isolate CIA-1) (CAV).
OC   Viruses; ssDNA viruses; Circoviridae; Circovirus.
OX   NCBI_TaxID=73478;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97126092; PubMed=8971016;
RA   Renshaw R.W., Soine C., Weinkle T., O'Connell P.H., Ohashi K.,
RA   Watson S., Lucio B., Harrington S., Schat K.A.;
RT   "A hypervariable region in VP1 of chicken infectious anemia virus
RT   mediates rate of spread and cell tropism in tissue culture.";
RL   J. Virol. 70:8872-8878(1996).
RN   [2]
RP   REVISION TO 50.
RA   Renshaw R.W.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY ACT AS TRANSCRIPTIONAL REGULATOR. INDUCES APOPTOSIS
CC       IN INFECTED CELLS. ELEMENT OF INFECTIOUS REPLICATION CYCLE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEUS OF INFECTED CELLS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L14767; AAD09423.1; -.
KW   Apoptosis; Nuclear protein.
SQ   SEQUENCE   121 AA;  13245 MW;  D2AAB39C2BD61A3E CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 121;
  Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     700 TLSLCG 705
           ||||||
Db      43 TLSLCG 48


RESULT 97
VCL1_PEA
ID   VCL1_PEA       STANDARD;      PRT;   124 AA.
AC   P02856;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vicilin, 14 kDa component.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=cv. Feltham First;
RA   Hirano H., Gatehouse J.A., Boulter D.;
RT   "The complete amino acid sequence of a subunit of the vicilin seed
RT   storage protein of pea (Pisum sativum L.).";
RL   FEBS Lett. 145:99-102(1982).
CC   -!- FUNCTION: SEED STORAGE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC       BODIES.
CC   -!- MISCELLANEOUS: THE MICROSEQUENCING TECHNIQUE DID NOT DISTINGUISH
CC       LEU FROM ILE; HOWEVER, RESIDUES AT POSITIONS 6, 23, 44, 86, 87,
CC       97, AND 112 ARE MOST LIKELY TO BE LEU AS THESE ARE POINTS OF
CC       CHYMOTRYPTIC CLEAVAGE.
CC   -!- SIMILARITY: BELONGS TO THE 7S SEED STORAGE PROTEIN FAMILY.
DR   PIR; A03346; FWPMV4.
DR   HSSP; P50477; 1CAU.
DR   InterPro; IPR001113; Seedstore_7s.
DR   Pfam; PF00546; Seedstore_7s; 1.
DR   Pfam; PF02808; Seedstore_7s_C; 1.
KW   Seed storage protein.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      18     18       R -> E.
FT   VARIANT      39     39       S -> N.
FT   VARIANT      47     47       E -> S.
FT   VARIANT      47     47       E -> R.
FT   VARIANT      48     48       D -> N.
SQ   SEQUENCE   124 AA;  14039 MW;  64D30A819E9CD6D3 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     828 GSLLLP 833
           ||||||
Db      84 GSLLLP 89


RESULT 98
ACPS_THETN
ID   ACPS_THETN     STANDARD;      PRT;   127 AA.
AC   Q8R857;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE   (4'-phosphopantetheinyl transferase acpS).
GN   ACPS OR TTE2171.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4 / JCM 11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
```

```
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
CC   -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC       A to a Ser of acyl-carrier protein (By similarity).
CC   -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC       3',5'-bisphosphate + holo-[acyl-carrier protein].
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE P-PANT TRANSFERASE SUPERFAMILY. ACPS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE013164; AAM25330.1; -.
KW   Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
KW   Complete proteome.
FT   METAL         8      8       MAGNESIUM (BY SIMILARITY).
FT   METAL        56     56       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   127 AA;  14386 MW;  8D975D2036A6BDD1 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 127;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     118 LGTGIR 123
           ||||||
Db      62 LGTGIR 67


RESULT 99
CD59_AOTTR
ID   CD59_AOTTR     STANDARD;      PRT;   128 AA.
AC   P51447;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CD59 glycoprotein precursor (Membrane attack complex inhibition
DE   factor) (MACIF) (MAC-inhibitory protein) (MAC-IP) (Protectin).
GN   CD59.
OS   Aotus trivirgatus (Night monkey) (Douroucouli).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX   NCBI_TaxID=9505;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95104908; PubMed=7528724;
RA   Fodor W.L., Rollins S.A., Bianco-Caron S., Burton W.V.,
RA   Guilmette E.R., Rother R.P., Zavoico G.B., Squinto S.P.;
RT   "Primate terminal complement inhibitor homologues of human CD59.";
RL   Immunogenetics 41:51-51(1995).
CC   -!- FUNCTION: POTENT INHIBITOR OF THE COMPLEMENT MEMBRANE ATTACK
CC       COMPLEX (MAC) ACTION. ACTS AT OR AFTER THE C5B-8 STAGE OF MAC
CC       ASSEMBLY.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: CONTAINS 1 UPAR/LY6 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L22861; AAA35372.1; -.
DR   HSSP; P13987; 1CDS.
DR   InterPro; IPR001526; LY6_UPAR.
DR   InterPro; IPR003632; Ly-6_CD59.
DR   Pfam; PF00021; UPAR_LY6; 1.
DR   ProDom; PD003128; Ly-6_CD59; 1.
DR   SMART; SM00134; LU; 1.
DR   PROSITE; PS00983; LY6_UPAR; 1.
KW   Antigen; Glycoprotein; GPI-anchor; Signal.
FT   SIGNAL        1     25       BY SIMILARITY.
FT   CHAIN        26    102       CD59 GLYCOPROTEIN.
FT   PROPEP      103    128       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   DOMAIN       26    108       UPAR/LY6.
FT   DISULFID     28     51       BY SIMILARITY.
FT   DISULFID     31     38       BY SIMILARITY.
FT   DISULFID     44     64       BY SIMILARITY.
FT   DISULFID     70     88       BY SIMILARITY.
FT   DISULFID     89     94       BY SIMILARITY.
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       102    102       GPI-ANCHOR (BY SIMILARITY).
SQ   SEQUENCE   128 AA;  14200 MW;  62D219B95589E55B CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     657 DCTFSR 662
           ||||||
Db      69 DCTFSR 74


RESULT 100
VL05_VACCV
ID   VL05_VACCV     STANDARD;      PRT;   128 AA.
AC   P07615;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein L5 (Protein F6).
GN   L5R OR F6.
OS   Vaccinia virus (strain WR), and
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254, 10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WR;
RX   MEDLINE=85215527; PubMed=2987815;
RA   Plucienniczak A., Schroeder E., Zettlmeissl G., Streeck R.E.;
RT   "Nucleotide sequence of a cluster of early and late genes in a
RT   conserved segment of the vaccinia virus genome.";
RL   Nucleic Acids Res. 13:985-998(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Copenhagen;
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [3]
RP   COMPLETE GENOME.
RC   STRAIN=Copenhagen;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES L5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01978; CAA26014.1; -.
DR   EMBL; M35027; AAA48080.1; -.
DR   PIR; E23092; QQVZF6.
DR   PIR; C42513; C42513.
SQ   SEQUENCE   128 AA;  15044 MW;  3F2D9A60A834ECE5 CRC64;

  Query Match             0.6%;  Score 6;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     757 SQPVSL 762
           ||||||
Db     117 SQPVSL 122


Search completed: April 22, 2003, 15:34:48
Job time : 63 secs
```